Metagenomics unveils the attributes of the alginolytic guilds of sediments from four distant cold coastal environments

Marina N Matos[1], Mariana Lozada[1], Luciano E Anselmino[1], Matías A Musumeci[1], Bernard Henrissat[2,3,4], Janet K Jansson[5], Walter P Mac Cormack[6,7], JoLynn Carroll[8,9], Sara Sjöling[10], Leif Lundgren[11] and Hebe M Dionisi[1]*

[1]Laboratorio de Microbiología Ambiental, Centro para el Estudio de Sistemas Marinos (CESIMAR, CONICET), Puerto Madryn, U9120ACD, Chubut, Argentina

[2]Architecture et Fonction des Macromolécules Biologiques, CNRS, Aix-Marseille Université, 13288 Marseille, France

[3]INRA, USC 1408 AFMB, F-13288 Marseille, France

[4]Department of Biological Sciences, King Abdulaziz University, Jeddah, 21589, Saudi Arabia

[5]Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA 99352, USA

[6]Instituto Antártico Argentino, Ciudad Autónoma de Buenos Aires, C1064ABR, Argentina

[7]Instituto Nanobiotec, CONICET- Universidad de Buenos Aires, Ciudad Autónoma de Buenos Aires, C1113AAC, Argentina

[8]Akvaplan-niva, Fram – High North Research Centre for Climate and the Environment, NO-9296 Tromsø, Norway

[9]CAGE - Centre for Arctic Gas Hydrate, Environment and Climate, UiT The Arctic University of Norway, N-9037 Tromsø, Norway

[10]School of Natural Sciences and Environmental Studies, Södertörn University, 141 89 Huddinge, Sweden

[11]Stockholm University, SE-106 91 Stockholm, Sweden

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

Running title: Alginolytic guilds from cold sediments

***Correspondenc**e: Hebe M. Dionisi, Laboratorio de Microbiología Ambiental, Centro para el Estudio de Sistemas Marinos (CESIMAR, CONICET), Blvd. Brown 2915, U9120ACD, Puerto Madryn, Chubut, Argentina. Tel. +54-280-4883184 extension 1251; Fax +54-280-4883543. e-mail: hdionisi@cenpat-conicet.gob.ar.

Originality-Significance Statement. The alginate degradation potential of high-latitude coastal sediments and the gene pool related to this process, which has a key role in brown algae biomass decomposition, were comprehensively analyzed using a metagenomic approach.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

Summary

Alginates are abundant polysaccharides in brown algae that constitute an important energy source for marine heterotrophic bacteria. Despite the key role of alginate degradation processes in the marine carbon cycle, little information is available on the bacterial populations involved in these processes. The aim of this work was to gain a better understanding of alginate utilization capabilities in cold coastal environments. Sediment metagenomes from four high-latitude regions of both Hemispheres were interrogated for alginate lyase gene homolog sequences and their genomic context. Sediments contained highly abundant and diverse bacterial assemblages with alginolytic potential, including members of Bacteroidetes and Proteobacteria, as well as several poorly characterized taxa. The microbial communities in Arctic and Antarctic sediments exhibited the most similar alginolytic profiles, whereas brackish sediments showed distinct structures with a higher proportion of novel genes. Examination of the gene neighborhood of the alginate lyase homologs revealed distinct patterns depending on the potential lineage of the scaffolds, with evidence of evolutionary relationships among alginolytic gene clusters from Bacteroidetes and Proteobacteria. This information is relevant for understanding carbon fluxes in cold coastal environments and provides valuable information for the development of biotechnological applications from brown algae biomass.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

Wiley-Blackwell and Society for Applied Microbiology

Introduction

Marine macroalgae present high photosynthesis and productivity rates even in highly challenging conditions such as those prevalent in polar environments, and they can contribute to long-term sedimentary carbon sequestration (Chung *et al.*, 2011; Wiencke and Amsler, 2012). Both macroalgae and the microorganisms specialized in the degradation of algal biomass are therefore key components of the carbon cycle in these rapidly changing environments (Arnosti *et al.*, 2013; Quartino *et al.*, 2013; Krause-Jensen and Duarte, 2014; Berlemont and Martiny, 2015). For instance, microbial communities from Arctic sediments are able to depolymerize a broad range of algal polysaccharides at high rates, acting as a final filter before carbon sequestration (Arnosti, 2008; Arnosti, 2014). Despite the key role of polysaccharide utilization for carbon fluxes in coastal ecosystems, the microbial populations participating in the degradation of algal polysaccharides and the mechanisms that they use are still poorly understood (Thomas *et al.*, 2012; Hehemann *et al.*, 2014; Wietz *et al.*, 2015). Metagenomics can provide unique insights into the polysaccharide utilization capabilities of these environments. Furthermore, this approach presents an opportunity for bioprospecting genes associated with polysaccharide utilization processes, which can be used for the development of biotechnological products from macroalgal biomass (Wargacki *et al.*, 2012; Lozada and Dionisi, 2015; Zhu and Yin, 2015).

Alginates are major components of the cell wall of brown algae, constituting up to 40% of their dry weight (Donati and Paoletti, 2009). They are linear polysaccharides composed of 1,4-linked β-D-mannuronate (M) and α-L-guluronate (G), arranged in Poly-M, Poly-G or Poly-MG blocks with 20-30 units each (Donati and Paoletti, 2009). The M/G ratio and the distribution of these blocks depend not only on the algal species, but also on the season, geographic location and type and age of the algal tissues (Kloareg and Quatrano, 1988; Truus *et al.*, 2001). The first step for the utilization of alginate as carbon source by marine bacteria is the depolymerization of these polysaccharides catalyzed by alginate lyases (AL; EC 4.2.2.3 and 4.2.2.11) through β-

elimination (Zhu and Yin, 2015). In the model mechanism proposed for the marine flavobacterium *Gramella forsetii* KT0803, endolytic AL enzymes attached to the outer membrane catalyze the depolymerization of alginates into oligoalginates (Kabisch *et al.*, 2014). The oligomers are then transferred through the outer membrane with the participation of TonB-dependent receptors (TBDR) and sugar-binding proteins. In the periplasm, the unsaturated oligosaccharides are further degraded to monomers by oligoalginate lyases and transported through the inner membrane, and in the cytoplasm a series of enzymes catalyze the conversion of the transported unsaturated monosaccharides into the metabolic intermediates glyceraldehyde-3-phosphate and pyruvate (Kabisch *et al.*, 2014). A similar process has been recently reported for the gammaproteobacterium *Saccharophagus degradans* 2-40 (Takagi *et al.*, 2016).

AL are remarkably diverse and they are currently classified into seven different polysaccharide lyase (PL) families, according to the classification scheme of the manually curated CAZy database (Carbohydrate-Active enZymes; Lombard *et al.*, 2014). The true diversity of these enzymes is only starting to emerge, as numerous AL sequences still remain unclassified or are yet to be included in the database (Garron and Cygler, 2010). Furthermore, the majority of the AL sequences known to date were identified in cultured bacteria, mostly belonging to the Flavobacteriia and Gammaproteobacteria classes (Zhu and Yin, 2015). Marine strains specialized in the utilization of alginates as substrate for growth contain several AL genes belonging to different PL families within their genomes (Thomas *et al.*, 2012; Mann *et al.*, 2013; Kabisch *et al.*, 2014; Neumann *et al.*, 2015, Takagi *et al.*, 2016), and the number of AL genes per genome has been shown to be related to its alginate degrading capability (Neumann *et al.*, 2015). Recently, studies that used culture-independent approaches started to shed light into the taxa potentially involved in alginate degradation in seawater (Wietz *et al*., 2015) and alginate gel particles (Mitulla *et al*., in press), the biogeographic distribution of AL sequences from *Alteromonas macleodii* ecotypes (Neumann *et al*., 2015), as well as features of alginate-

degrading consortia growing in anaerobic conditions (Seon *et al.*, 2014; Kita *et al.*, 2016). However, the alginate utilization potential of microbial communities from coastal environments and the genes involved in these processes still remain largely unknown (Neumann *et al.*, 2015).

Brown macroalgae are an important ecosystem component in cold coastal environments (Wiencke and Amsler, 2012). Part of their dead biomass is transferred to surface sediments, where heterotrophic bacteria participate in its decomposition (Hardison *et al.*, 2010). Due to the high alginate content in these macroalgae, we hypothesized that bacterial populations with the ability to use these polysaccharides as carbon source would be abundant in the microbial communities from sediments of these environments. The aim of this work was to increase our understanding of alginate utilization processes in cold coastal environments. Using a multi-level metagenomic approach, we characterized putative AL sequences and their genomic context in sediments from four distant high-latitude environments. The analysis of the gene pool related to alginate utilization suggested that these populations are both remarkably abundant and diverse, and include members of poorly described taxa. In addition, this work sheds light into evolutionary relationships of the alginolytic capabilities across taxa.

Results and Discussion

Study sites

Four high-latitude coastal environments were analyzed in this study (referred as sampling regions, Fig. 1 and Fig. S1). Polar habitats included Advent fjord (Spitsbergen Island, Svalbard Archipelago, Norway), and Potter Cove (25 de Mayo [King George] Island, South Shetland Islands, Antarctica). Subpolar habitats included Ushuaia Bay (Tierra del Fuego Island, Argentina) and Baltic Sea, the only brackish environment analyzed in this study (Table S1). Although many algae species present a bipolar distribution, the macroalgae communities are

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

and Amsler, 2012). The four sampling regions present some level of anthropogenic impact, such as eutrophication and/or chronic contamination with toxic pollutants. More information on these topics can be found in Supporting Information, General Features of the Study Sites. The nested sampling approach included triplicate subtidal sediment samples obtained at two sites in each of the four regions, and 23 of these samples produced the metagenomic dataset.

Abundance of genes encoding putative PL enzymes

Although Pfam domains are not suitable for the classification of CAZymes into specific PL families (Lombard *et al.*, 2014), the abundance of sequences containing domains related to PL enzymes in the gene pool of the metagenomes (relative to the abundance of single-copy genes) can provide an estimation of the prevalence of genes encoding AL and other PL enzymes in these environments. Three of the four Pfam domains found in AL enzymes, PF05426 (Alginate lyase), PF07940 (Heparinase II/III-like protein) and PF08787 (Alginate lyase 2), were abundant in the metagenomes, while PF14592 (Chondroitinase B) was present at lower abundances (Fig. 2). Two of these domains, PF07940 and PF08787, showed significant differences in relative gene abundance among sampling regions (Kruskal-Wallis test, $p = 0.004$ for both domains), which could reflect distinct structures of the alginolytic guilds or differences in the set of AL genes present in members of these guilds. The largest differences were found in the domain PF08787, identified in sequences of the PL7 and PL18 families, both encompassing only AL enzymes (Garron and Cygler, 2010). The relative abundance of this domain varied from 0.12 ± 0.01 copies in Baltic Sea sediments to 0.99 ± 0.28 copies in Antarctic sediments, per single-copy gene (Fig. 2). Pearson correlation analysis revealed a positive correlation between the relative abundance of sequences containing a PF08787 domain and salinity ($r = 0.67$, $p < 0.001$, n = 23). Although less marked, differences in the abundance of this domain were also observed between the two sampling sites from Ushuaia Bay (0.80 ± 0.21 copies in MC [samples ARG01-ARG03] and 0.33 ± 0.09 copies in OR [ARG04-ARG06], per single-copy

gene). Small differences in salinity were also detected at these sites, probably due to the influence of a freshwater runoff next to OR site (Table S1). A positive correlation was also found between the relative abundance of PF08787 domain and salinity when considering only the samples from Ushuaia Bay (Pearson correlation coefficient $r = 0.85$, $p = 0.032$, n = 6). These results suggest that salinity could be a factor influencing the abundance of AL genes from the PL7 and/or PL18 families in the sediment microbial communities. A previous study also detected a correlation between gene patterns (*alkB*) and salinity in the metagenomes from samples ARG01-06 (Guibert *et al*., 2016). It is important to notice, however, that other environmental factors, such as pollution, could also be influencing the observed gene distribution patters (Supporting Information, General Features of the Study Sites).

The overall abundance of AL genes in the metagenomes was estimated considering the domains PF05426, PF08787 and PF14592. As the latter is also present in chondroitinases, these values can be slightly overestimated. AL gene abundances were significantly higher in the three marine environments than in the Baltic Sea ($p = 0.036$ for all three post-hoc tests, Bonferroni corrected). Remarkably, the overall abundance of AL genes relative to single-copy genes in marine sediments ranged from 1.12 ± 0.18 copies in Svalbard to 1.78 ± 0.46 copies in Antarctica. The high potential for alginate depolymerization in marine sediments could be due to the selection of bacterial populations with this trait as a result of a high brown algae biomass production (Quartino and De Zaixso, 2008; Liuzzi *et al.*, 2011; Wiencke and Amsler, 2012). Sediments from the Arctic archipelago Svalbard were previously shown to depolymerize alginates at high rates, even under anaerobic conditions (Arnosti, 2008, Teske *et al.*, 2011). Although similar extracellular enzymatic activities have not been determined in Antarctic sediments, the high abundance of AL genes in this environment suggests a high potential for alginate depolymerization. Despite this potential, algal biomass decomposition is a slow process at the sampling location in Potter Cove, where insufficient mechanical break-down and low temperatures could be limiting the macroalgae degradation rate (Quartino *et al.*, 2008; Quartino

et al., 2015). The overall abundance of AL genes in sediments from the Baltic Sea was 0.66 ± 0.04 copies per single-copy gene, suggesting a lesser prevalence of microorganisms with alginolytic potential in sediments of this brackish environment, possibly due to a lower biomass of brown algae (Alexandridis *et al.*, 2012; General Features of the Study Sites in Supporting Information). Homologues of AL genes could remain undetected in this analysis if they are divergent from those used to define the Pfam domains.

A domain found in pectin-degrading enzymes (PF09492) was highly abundant in the sediments and clustered with the four domains from AL enzymes (top cluster, Fig. 2). Reports of pectinolytic marine microorganisms in the literature are scarce (Tuyen *et al.*, 2001), but the presence of organic carbon from terrestrial origin is common in coastal sediments (Wysocki *et al.*, 2008; Sikes *et al.*, 2009) and the ability to utilize plant polysaccharides might represent an advantage for microorganisms inhabiting near-shore environments. Polar and subpolar sediments could therefore represent promising sources for bioprospecting plant biomass-degrading enzymes, as previously suggested for other marine habitats (Vigneron *et al.*, 2014). Other protein families containing pectate lyases, on the other hand, were less abundant and formed a second cluster with domains from other PL enzymes (bottom cluster, Fig. 2).

Ordination and clustering of metagenomes based on shared alginate lyase homolog (ALH) sequences

We identified 2,705 sequences (≥100 amino acids) homologous to AL genes in the assembled metagenomes. In spite of the low level of read assembly in the metagenomes, 13.3% of the dataset were full-length sequences (for a full description of the dataset, see Characteristics of the Alginate Lyase Homolog Sequence Dataset in Supporting Information). As only seven sequences were identified in each of the metagenomes ARG04 and ARG06, these samples were excluded from further analyses. We evaluated the diversity of the 2,961 remaining ALH sequences by grouping them into operational protein units (OPUs) defined at 80% identity at the

amino acid level. We used a greedy incremental algorithm, which selects the longer sequences as seeds for OPU building (Fu *et al.*, 2012), to reduce the risk of overestimating the number of OPUs in the metagenomic dataset. This analysis identified 690 OPUs, suggesting a high diversity of microbial populations with the potential to depolymerize alginates in these sediments. On ordination plots, the samples retrieved from each region clustered together (Fig. 3), indicating the presence of distinct alginolytic guilds in the different coastal environments. Temperature and salinity correlated with the major ordination axes (Pearson's product-moment correlation coefficients r = 0.77 and 0.93, respectively; p = 0.001 for both correlations), suggesting that these environmental factors could play a role in the structuring of the alginolytic guilds. Similarly, the distribution pattern of sequences encoding chitinases, another enzyme targeting polysaccharides, was found to be strongly influenced by salinity, among other factors (Beier *et al*., 2010). Salinity is a key driver of the structuring of microbial communities, probably through selection processes (Herlemann *et al*., 2011; Thureborn *et al*., 2013). As this study did not include a salinity gradient, however, it is not possible to establish a direct link between the distribution of AL genes and salinity.

Ordination and its superimposed hierarchical clustering analysis showed that the ALH gene pools from both polar environments (Antarctica and Svalbard) were the most similar (Fig. 3). These pools were further related to those from Ushuaia Bay, a subantarctic marine environment. By contrast, the samples from the Baltic Sea, a brackish environment, clustered separately. These datasets shared only 3% and 3.2% of the OPUs with samples from Svalbard and Antarctica, respectively. This result is in agreement with the unique evolutionary lineages found in the Baltic Sea, due to its recent geological history, distinctive environmental conditions and geographic isolation, including both marine and freshwater microorganisms (Johannesson and André, 2006; Herlemann *et al.*, 2011). Despite the differences found in the ALH sequence datasets among sampling regions, four OPUs (representing 7.5% of the dataset) included sequences identified in metagenomes from the four sampling regions. The remarkable

similarities in this subset of ALH sequences, despite large geographic distances and differences in environmental conditions (Hanson *et al.*, 2012; Zinger *et al.*, 2014), suggests that some alginolytic populations from cold coastal environments could have a widespread geographic distribution. However, the possibility of recent horizontal gene transfer events cannot be excluded, and therefore these closely-related genes could be hosted by unrelated taxa.

Classification of the metagenomic ALH sequences

We assessed the level of relatedness between each metagenomic sequence and its closest CAZyme sequence (Lombard *et al.*, 2014), by comparing the E-values obtained using standalone blast (Altschul *et al.*, 1990) with all PL sequences as a custom database (Fig. 4A). Baltic Sea sediments contained a lower proportion of sequences with highly significant E-values ($< 1e^{-50}$) when compared with those from Antarctica and Svalbard (Kruskal-Wallis test and post-hoc analysis, Bonferroni corrected $p = 0.036$), suggesting the presence of a larger proportion of novel ALH sequences in brackish sediments. Similarly, when the metagenomic sequences were classified using internal CAZy tools (Cantarel *et al.*, 2009), a third of the sequences from the Baltic Sea samples and from ARG05 could not be classified, as they were too distantly related to CAZymes for a reliable assignment (Fig. 4B). Overall, 13% of the ALH sequences could not be classified. These sequences clustered into 184 OPUs (80% identity level), evidencing the high diversity of novel ALH sequences contained in these metagenomes. On the other hand, the majority of the ALH sequences were classified as belonging to the PL17 (30.5%), PL7 (28.2%) and PL6 (22.1%) families (Fig. 4B). AL genes from these families have previously been identified in the genomes of several marine alginolytic bacteria, often in more than one copy (Thomas *et al.*, 2012, Kabisch *et al.*, 2014; Takagi *et al.*, 2016). In the model suggested for the flavobacterium *Gramella forsetii* KT0803, PL7 and PL6 endo-acting AL are located on the outer membrane and PL17 oligoalginate lyases are within the periplasm (Kabisch *et al.*, 2014).

Interestingly, the proportion of sequences from the PL7 family was lower in the metagenomes

from Baltic Sea sediments and in the sample ARG05. In contrast, samples ARG01-ARG03, retrieved at a 500 m distance from sample ARG05, showed a distribution of PL families more similar to those from Svalbard and Antarctica. These results are in agreement with the lower abundance of its characteristic domain (PF08787) in the total metagenomes from Baltic Sea sediments and ARG04-06 samples, shown in Fig. 2.

Analysis of full-length ALH sequences

We performed a phylogenetic analysis based on the catalytic modules of classified full-length ALH sequences, sequences from the CAZy database and homologous sequences identified in bacterial genomes. Within each PL family, the metagenomic sequences were phylogenetically diverse, as they were widely distributed across different clusters of the trees. Fig. S3 shows the phylogenetic tree of the PL6 family, subfamily 1, and Fig. 5 shows two clusters selected from this tree. Similar clustering patterns were observed in other PL families (data not shown). Several clusters included sequences identified in the metagenomes from both polar environments, sharing high identity values (Fig. S3 and Fig. 5A). Often, these ALH sequences were related to AL genes described in members of the Bacteroidetes (e.g. *Winogradskyella*, *Polaribacter*, *Gramella*) or Proteobacteria (*Alteromonas*, *Pseudoalteromonas*). By contrast, ALH sequences from brackish sediments clustered separately, not only from other metagenomic sequences but also from sequences from bacterial strains (Fig. S3 and Fig. 5B). These results are in accordance with the ordination analysis of the full dataset (Fig. 3).

The gene neighborhoods of the ALH sequences often included additional ALH sequences identified in the dataset (Fig. 5). In addition, they contained other sequences potentially related to alginate utilization, previously described in alginolytic operons from flavobacteria (Thomas *et al.*, 2012; Kabisch *et al.*, 2014). Hypothetical proteins were also often found in these clusters. For instance, two hypothetical proteins located downstream of a putative

susC gene (which encodes an outer membrane protein associated with polysaccharide utilization, Tang *et al.*, 2012) in scaffolds from Baltic Sea sediments were also found in the genome of the Bacteroidetes *Rhodothermus marinus* DSM 4252 (Fig. 5B) and in the composite genome from an uncultured bacterium belonging to the Ignavibacteriae phylum (NCBI taxon ID 795747, data not shown). The gene order conservation of these three sequences in members of different phyla suggests that these hypothetical proteins could be related to alginate utilization.

A high level of gene order conservation was often found in scaffolds from the same sampling region containing very similar ALH sequences (Fig. 5). Partial shared synteny was also observed among scaffolds containing similar ALH sequences identified in distant sampling regions (Antarctica and Svalbard, Fig. 5A), providing additional evidence that bacterial populations with alginolytic potential could have a broad geographic distribution. Gene order conservation with genomes from bacterial isolates was less frequent, although in some cases the same sequences were found in the gene neighborhood of the ALH sequences, but in a different order (Fig. 5B). An exception was a series of scaffolds from Antarctica, which shared a remarkably conserved synteny with a section of the genome of *Psychromonas arctica* DSM 14288 (Fig. S4). The longest scaffold (~35 Kb) shared 80.8% identity at the nucleotide level with the genome of this gammaproteobacterium, which was isolated from seawater near Svalbard (Groudieva *et al.*, 2003). Although ALH sequences highly similar to those from *P. arctica* DSM 14288 could not be detected in the assembled metagenomes from the other three regions analyzed in this study, sequences sharing 99 - 100% identity with the 16S rRNA genes from this organism were identified in the metagenomes from the four regions (data not shown). Strains belonging to the genus *Psychromonas* have been isolated from several cold marine environments, and it has been proposed that ocean circulation linked to the sinking of cooled seawater in polar regions might facilitate the dispersion of this cold adapted microorganism (Nogi *et al.*, 2002).

Among the 359 full-length ALH sequences identified in the metagenomes, 50 could not be assigned to a CAZyme class or family. These full-length unclassified sequences formed 24 clusters using an 80% amino acid identity cut-off. In order to assess if these unclassified sequences could code for AL enzymes, we further analyzed a representative sequence from each cluster (Table S3). The majority of these sequences contained conserved Pfam domains characteristic of AL enzymes (mostly PF05426). Furthermore, often their closest matches in blastp searches against the non redundant protein NCBI database were AL sequences from bacterial strains or hypothetical proteins with similar Pfam domains, the majority sharing low to moderate identity values (Table S3). Four unclassified sequences containing a PF05426 domain were selected for *in silico* structure prediction. The three-dimensional modeling of these ALH sequences suggested that they could present a PL5-type fold, which is characteristic of sequences containing this Pfam domain (more information can be found in the Supporting Information, Three-dimensional Models of ALH Sequences; Fig. S5 and S6). Interestingly, in the structural models of the metagenomic sequences NOR15_100118034 and SWE02_100064912, the residues involved in the acid/base mechanism in AL enzymes and some of the amino acids involved in the interaction with the substrate were not identified in the equivalent position (Table S4). In contrast, the predicted structures of two consecutive unclassified ALH sequences from scaffold SWE02_10000683 (Fig. 5B and Fig. S6) showed a higher conservation of key residues (Table S4). These novel metagenomic sequences constitute promising targets for heterologous expression and enzymatic characterization.

Genomic context of ALH sequences in different taxonomic groups

Very limited information is available on the diversity and taxonomic identity of bacterial populations with alginolytic potential from marine environments (Wietz *et al.*, 2015). We analyzed the taxonomic assignment of scaffolds ≥ 4 kb in length that included ALH sequences, as longer sequences are able to provide more reliable taxonomic information (Patil *et al.*, 2012).

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

Although only part of the alginolytic guild was represented in this analysis, this set included 206 scaffolds covering a total of 2.2 Mb, and contained 319 ALH sequences (Table S5). We used two methods for the taxonomic binning, a sequence-similarity based method (IMG/M pipeline; Huntemann *et al.*, 2016) and a composition-based method (Patil *et al.*, 2012). The assignments were mostly in agreement at high taxonomic levels (90% at the phylum or superphylum level, Table S5). In fragments from the Baltic Sea, however, almost a third of the scaffolds were classified into different phyla with the two methods and a higher proportion of the scaffolds could not be classified at the phylum level, suggesting that these microorganisms are divergent from those represented in the databases. In the full dataset, twelve percent of the scaffolds could only be affiliated to Bacteria or could not be assigned using the similarity-based method (Table S5). Most of these scaffolds were classified into 10 different phyla using composition-based taxonomic binning, including Bacteroidetes and Proteobacteria, as well as poorly characterized taxa such as Ignavibacteriae, Cloacimonetes, Verrucomicrobia and Planctomycetes (Table S5 and Fig. S7). These phyla were relatively abundant in the sediment microbial communities (Fig. S8A).

Approximately half of the scaffolds binned within the Bacteroidetes phylum using both approaches, the majority of them within the Flavobacteriia class (Table S5 and Fig. S7). Bacteroidetes represented the second most abundant phylum in the sediment microbial communities (Fig. S8A), and the Flavobacteriia class was predominant within this phylum (>90%, Fig. S8B). On the other hand, between a quarter (sequence-similarity based method) and a third (composition-based method) of the scaffolds were assigned to the Proteobacteria phylum (Table S5 and Fig. S7). The majority of these scaffolds were identified in sediments from Antarctica, where two third of these scaffolds binned within the Gammaproteobacteria class and almost half were assigned to the order Alteromonadales. Members of the Gammaproteobacteria class were most abundant in microbial communities from sediments of Antarctica (Fig. S8C), and half of the members of this class belonged to the order Alteromonadales (data not shown).

These results suggest that members of the Bacteroidetes and Proteobacteria phyla could have an important role in the alginolytic guilds of cold coastal sediments. Gene clusters associated with alginate utilization have been identified in the genomes of several members of these phyla, mostly within the Flavobacteriia and Gammaproteobacteria classes (Thomas *et al.*, 2012). In a recent study that analyzed the structure and metabolic potential of an alginate-degrading anaerobic consortium enriched from coastal sediments, members of the Bacteroidetes and Firmicutes phyla were the dominant members, while populations belonging to the Gammaproteobacteria class were less abundant (Kita *et al.*, 2016). Conversely, only members of the Gammaproteobacteria class (in particular from the order Alteromonadales) were enriched in seawater microcosms after incubation with sodium alginate (Wietz *et al.*, 2015). In a more recent study, the same research group found a higher diversity of taxa colonizing and degrading alginate-gel particles in microcosms with seawater, including members of the Roseobacter clade (Alphaproteobacteria), as well as Cryomorphaceae (Flavobacteriia class) and Saprospiraceae (Sphingobacteriia) families (Mitulla *et al.*, in press). More studies are needed to identify the role of the environmental matrix on the structuring of the alginolytic guild in coastal marine environments. The capability for migration and surface attachment of certain groups could favor their colonization of certain niches such as marine snow or sediments (Cordero and Datta, 2016).

We further analyzed if the genomic context of the ALH sequences from the selected scaffolds varied among the different taxa, based on the potential function of the sequences (significant hits of Pfam domains; Finn *et al.*, 2013) located in codirectional gene clusters. Divergently transcribed genes, even if they were in some cases related to alginate metabolism (for instance *gntR*-like genes found adjacent but in opposite orientation), were not included in this analysis. Clustering analysis based on Bray-Curtis similarity index revealed a high similarity in the genomic context of ALH sequences in scaffolds from Antarctic and Svalbard metagenomes assigned to the Bacteroidetes phylum (Fig. S9). Scaffolds that binned within the

Bacteroidetes and Proteobacteria phyla from Antarctica, Svalbard and Baltic Sea metagenomes formed a larger cluster. These results support the proposed hypothesis of a common phylogenetic origin for these gene clusters, with Proteobacteria acquiring alginate utilization systems by several independent horizontal gene transfer events from ancestral marine flavobacteria (Thomas *et al.*, 2012). Scaffolds from Svalbard and Baltic Sea metagenomes within the lineage Bacteria also clustered together, despite the large differences in GC content between these two sets of scaffolds (0.56 ± 0.03 and 0.35 ± 0.02, respectively). This clustering was mainly due to a high number of CDS containing the PF00884 domain (Sulfatase) in a third of these scaffolds. Many polysaccharides used by marine bacteria as substrate for growth are sulfated, and removing these groups from the polymer facilitates their degradation (Gerken *et al.*, 2013).

In some of the scaffolds that binned within the Flavobacteriia class, ALH sequences were located within large alginolytic gene clusters, while in other scaffolds the gene neighborhood of ALH was not associated with alginate utilization. The genome of flavobacterial strains isolated from various marine habitats, including cold coastal environments, contain both alginolytic gene clusters and isolated AL genes (Thomas *et al.*, 2012; Mann *et al.*, 2013; Kabisch *et al.*, 2014; Inoue *et al.*, 2014; Xing *et al.*, 2015). The alginolytic gene clusters from the metagenomic scaffolds resembled the highly conserved gene order previously reported in flavobacterial alginolytic clusters (Thomas *et al.*, 2012; Kabisch *et al.*, 2014). However, these clusters contained many deletions, insertions and duplications (Fig. S10), suggesting a higher diversity of gene organizations within this group than currently recognized. Seven Pfam domains related to components of alginolytic clusters were detected in the gene neighborhood of the ALH sequences in the Bacteroidetes scaffolds from Antarctica, Svalbard and Baltic Sea, besides the four domains associated with AL enzymes (PF05426, PF07940, PF08787 and PF14592, Fig. S9): (i) PF00392 (GntR) and PF07729 (FCD), identified in transcriptional regulators of the subfamily FadR of the GntR family (Rigali *et al.*, 2002); (ii) PF07690 (MFS_1), membrane

transporters belonging to the major facilitator superfamily (Reddy *et al.*, 2012); (iii) PF07883 (Cupin_2), proteins with a cupin-like domain; (iv) PF00106 (adh_short) and PF13561 (adh_short_C2), two of the domains found in members of the large superfamily of short-chain dehydrogenases/reductases (SDR; Persson and Kallberg, 2013); and (v) PF00294 (PfkB), family of carbohydrate kinases that include the enzyme KdgK involved in the cytoplasmic processing of alginate monomers (Cabrera *et al.*, 2010). Another set of Pfam domains were identified in Bacteroidetes scaffolds only from the metagenomes from Antarctica and Svalbard: (i) PF00593 (TonB_dep_Rec) and PF07715 (Plug) from SusC-like proteins (Tang *et al.*, 2012); (ii) PF07980 (SusD) and PF14322 (SusD-like_3) from SusD-like proteins (Tang *et al.*, 2012) and (iii) PF00801 (PKD) in PKD-domain containing proteins, potentially involved in protein–protein interactions in the outer membrane (Bauer *et al.*, 2006). These proteins have been proposed to participate in the transport of oligoalginates through the outer membrane (Kabisch *et al.*, 2014).

The gene organization of the scaffolds from Proteobacteria was highly variable (Table S5), and the ALH sequences were found in short clusters of codirectional genes. These results are in agreement with the limited shared synteny and short length reported in alginolytic gene clusters from strains belonging to the Proteobacteria phylum (Thomas *et al.*, 2012; Neumann *et al.*, 2015). Like in scaffolds assigned to the Bacteroidetes phylum, PF07883 (Cupin_2) was identified in gene clusters from all three regions, suggesting that these proteins could play a key role in alginate degradation in members of both Bacteroidetes and Proteobacteria phyla. In *Z. galactanivorans*, the gene containing this domain (*kdgF*) was found to be differentially expressed in the presence of alginate (Thomas *et al.*, 2012). In addition, the product of this gene has been detected in members of these two phyla growing on alginate: a cupin 2 conserved barrel protein was one of the 23 alginate-specific proteins of the gammaproteobacterium *S. degradans* 2-40 (Takagi *et al.*, 2016), and the KdgF protein was detected in the cytoplasm of the flavobacterium *G. forsetii* KT0803 (Kabisch *et al.*, 2014). Although the function of this protein is not known, it has been recently suggested that it could participate in the first step of the

cytoplasmic processing (linearization) of the unsaturated monouronic acids (Lee, 2014), which has been proposed to be spontaneous (Thomas *et al.*, 2012). Further studies are needed to determine the role of KdgF in alginate degradation. On the other hand, domains PF00106 (adh_short), PF7715 (Plug) and PF01081 (Aldolase), a family that includes the enzyme phospho-2-dehydro-3-deoxygluconate aldolase (KdgA), were only identified in scaffolds from both polar regions.

Overall, 17% of the sequences from the gene clusters did not have significant Pfam hits. Although some of these sequences could have been incorrectly predicted, other sequences could be related to alginate degradation. For instance, a protein of unknown function encoded in an alginate utilization cluster has been detected in the cytoplasm in *G. forsetii* KT0803 (Kabisch *et al.*, 2014). The expression and characterization of these sequences will be needed to elucidate their role in alginate utilization processes.

Concluding Remarks

The exploitation of metagenomic data to gain ecological and mechanistic insights into the metabolic potential of microbial communities is often hindered due to the highly fragmented nature of this information, and this limitation is even more critical for bioprospecting efforts (Lozada and Dionisi, 2015). In this work, we used different levels of resolution to analyze a highly-complex metagenomic dataset, capitalizing on the available information and accounting for the limitations imposed by the inefficient and uneven read assembly. Evidences of high abundance and diversity of bacteria with the potential to utilize alginates detected in sediments from geographically-distant cold coastal environments suggest that this process is not only highly relevant, but more complex than previously assumed. Although a fraction of the putative alginate lyase gene pool was shared among regions, a highly distinctive guild seems to be

play an important role in the structuring of these guilds. Besides members of the Bacteroidetes and Proteobacteria phyla, this study also revealed several largely unknown groups of bacteria with the potential to depolymerize alginates in polar and subpolar environments. In particular, sediments of the Baltic Sea represent an attractive target for further exploration, such as the isolation or single-cell genomic analysis of novel polysaccharide-degrading bacteria. The analysis of the genomic context of selected scaffolds confirmed at the community level evolutionary patterns proposed based on comparative genomic analyses (Thomas *et al.*, 2012). The differences in gene organization detected in alginolytic gene clusters suggest recent gene rearrangements in the genome of these microorganisms. These observations raise fundamental questions regarding the dispersal and adaptative capabilities of these microorganisms. Polar marine ecosystems are highly vulnerable to local and global anthropogenic impacts. The effects of climate-induced changes on microbial community structure and function, and their implications in ecosystem function, are still unknown. Identifying the microorganisms and understanding the mechanisms involved, as well as the environmental factors influencing alginate utilization processes, are a prerequisite for modeling the functioning of these rapidly changing ecosystems.

Experimental Procedures

Sample and metadata collection

Four high-latitude coastal environments were analyzed in this study (referred in the text as sampling regions, Fig 1, Table S1 and Fig. S1): (i) Advent Fjord, Spitsbergen, Svalbard Archipelago, Norway (NOR); (ii) Port Värtahamnen, Stockholm, Baltic Sea, Sweden (SWE); (iii) Ushuaia Bay, Tierra del Fuego Island, Argentina (ARG); (iv) Potter Cove, 25 de Mayo (King George) Island, Antarctica (ANT). Triplicate sediment samples (top 5 cm) were collected

were collected using cores, and environmental parameters, including temperature, depth and salinity, were determined *in situ* using CTD or multiparameter instruments (Table S1). Sediment samples were stored at -80ºC.

DNA extraction and metagenomic sequencing

Metagenomic DNA was extracted from the sediment samples as previously reported (Mackelprang *et al.*, 2011). Shotgun sequencing of metagenomic DNA was performed using Illumina HiSeq 2000 platform (2×150-bp paired end reads, one lane per sample), at the facilities of the Joint Genome Institute, USA. The dataset includes 23 metagenomes (Table S1), which were annotated using the IMG/M pipeline (Markowitz *et al.*, 2014). The number of reads and the assembly efficiency varied among samples, with <35% of the reads mapping the scaffolds (Table S1). The metagenomes are freely accessible at the IMG server (https://img.jgi.doe.gov/) under accession numbers 3300000118-3300000136, 3300000241-3300000243, and 3300000792.

Relative abundance of Polysaccharide Lyase (PL) sequences

Twelve Pfam domains contained in sequences from different PL families and 12 domains from single-copy genes that code for ribosomal proteins (Kunin *et al.*, 2008) were selected to estimate the relative abundance of PL sequences in the metagenomes (Table S2). The estimated copies of genes containing these domains in the total metagenomes were retrieved from the IMG/M system using the Abundance Profile Overview tool including both unassembled and assembled metagenomes, the latter corrected by read depth. The estimated copies of genes containing Pfam domains from PL were normalized by dividing by the estimated copies of each of the single-copy genes and the values obtained were averaged. Only the domains PF05426, PF08787 and PF14592 were used to estimate the overall abundance of AL genes in the metagenomes, to avoid counting twice members the PL17 family (which contains domains

PF07940 and PF5426). PF14592 can be detected in both AL and chondroitinases, and therefore the overall abundance could be slightly overestimated. To test for significant differences in gene abundance among sampling regions, Kruskal-Wallis tests were performed using SPSS v. 15 (α = 0.05 after Bonferroni correction for multiple comparisons). Clustering analyses based on Bray Curtis similarity were performed using R package *vegan* (https://cran.r-project.org/web/packages/vegan/index.html).

Identification and analysis of alginate lyase homolog (ALH) sequences

ALH sequences were identified in the assembled metagenomes using blastp (threshold E-value of $1e^{-5}$; Altschul *et al.*, 1990) with sequences representing the PL families containing AL sequences as query: GenBank accession numbers AEW23144, YP_004791784, YP_004791784, AFC88009, WP_007214684, NP_627697, YP_006746939, YP_002800319, YP_001347824, ACN56743, YP_003118157, YP_004738545, BAB03312, BAH79131, NP_624710, BAB19127, AGE48774, BAE81787, ADE10038, NP_357573, YP_528751, ACB87607 and YP_003656270. Protein coding sequences (CDS) with *alginate lyase* as product name were also retrieved. Duplicated sequences identified by both methods or by different query sequences were eliminated. Only sequences $\geq$ 100 amino acids were analyzed. ALH sequences were grouped into operational protein units (OPUs) defined at 80% identity at the amino acid level, using CD-HIT (Fu *et al.*, 2012). Ordination (NMDS) and clustering using Bray-Curtis similarity based on Wisconsin standardized OPU data, and environmental fitting to the metadata matrix were performed in R package *vegan*.

The level of divergence between each metagenomic sequence and the closest PL sequence of the CAZy database (http://www.cazy.org, as of February 2014) was assessed using standalone blast (Altschul *et al.*, 1990), through the E-value of the match. ALH sequences were classified at the family level with the CAZy pipeline, which uses a combination of blast and HMM tools and further subjected to manual curation (Lombard *et al.*, 2014). For the

phylogenetic analysis, only full-length metagenomic sequences were selected, which represented approximately 13% of the dataset. Phylogenetic trees were built for each PL family, including CAZymes and homologous sequences identified in bacterial genomes deposited at the IMG database (http://img.jgi.doe.gov/) using blastp searches. The PL module of the sequences was detected using dbCAN (Yin *et al.*, 2012), aligned using default parameters in ClustalX 2.1 (Larkin *et al.*, 2007), and manually curated and trimmed. Phylogenetic trees were built using the maximum likelihood algorithm in Mega (Tamura *et al.*, 2011) using WAG (PL6 SF1) or LG (PL17 SF2) substitution models, the most appropriate model in each case. The stability of tree topology was evaluated by bootstrap resampling with 500 replications.

Analysis of scaffolds containing ALH sequences

Phylogenetic lineage, lineage percentage, GC content, gene content and gene organization of the scaffolds were analyzed based on the information obtained from the functional annotation of the metagenomes (Markowitz *et al.*, 2014). Additionally, PhylopythiaS (Patil *et al.*, 2012) was used for the taxonomic binning of the scaffolds. For comparative analysis, gene organization of bacterial strains was retrieved from genomes deposited at the IMG database (Markowitz *et al.*, 2012). For the statistical analysis of gene neighborhood of ALH sequences, a Pfam search (Finn *et al.*, 2013) was performed for all the sequences located codirectional to the identified ALH sequences in scaffolds ≥4 kb. Clustering analysis of the gene content information discriminated by region and by lineage was performed based on Bray-Curtis similarity index calculated for the gene per sample matrix, standardized by Wisconsin double standardization, using R package *vegan*.

Microbial community structure

The analysis of the sediment microbial community structure was based on 16S rRNA gene information obtained from the metagenomes. The relative abundance of blast hits (90% identity) for bacterial sequences in the IMG/M system

using the tool Phylogenetic Distribution of 16S rRNA Genes for the unassembled and assembled metagenomes.

Acknowledgments

ML, MAM and HML are staff members from The National Research Council of Argentina (CONICET). MNM is a postdoctoral fellow from CONICET. The metagenomic dataset was generated at the Department of Energy-Joint Genome Institute (DOE-JGI) under the Community Sequencing Program (CSP proposal ID 328, project IDs 403959, 404206, 404777-404782, 404786, 404788- 404801). HD and ML were supported by grants from CONICET (Nº 112-200801-01736) and The National Agency for the Promotion of Science and Technology of Argentina (ANPCyT PICT2008 Nº 0468). WMC was supported by grants from the the University of Buenos Aires (UBA 2014-2017 20020130100569BA), the European Commission through the Marie Curie Action IRSES IMCONet (project N° 318718), the Argentinean Antarctic Institute and ANPCyT (PICTO 2010 N° 0124). JKJ was supported by the Pacific Northwest National Laboratory under Contract DE-AC05-76RLO1830. BH was supported by Agence Nationale de la Recherche, grant BIP:BIP (ANR-10-BINF-03-04). JC's research contribution is supported by the Research Council of Norway (grant No. 223259). We thank Krystle Chavarria for her technical support and Ricardo Vera and Horacio Ocariz for their help in sample collection. We would like to dedicate this work to our late colleagues Horacio Ocariz and Leif Lundgren, who left us too early.

The authors declare no conflict of interest.

References

Alexandridis, N., Oschlies, A., and Wahl, M. (2012). Modeling the effects of abiotic and biotic factors on the depth distribution of *Fucus vesiculosus* in the Baltic Sea. *Mar Ecol-Prog Ser* **463:** 59-72.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J Mol Biol* **215**: 403-410.

Arnosti, C. (2008). Functional differences between Arctic seawater and sedimentary microbial communities: contrasts in microbial hydrolysis of complex substrates. *FEMS Microbiol Ecol* **66:** 343-351.

Arnosti, C. (2014). Patterns of microbially driven carbon cycling in the ocean: links between extracellular enzymes and microbial communities. *Adv Oceanog* **2014:** ID 706082.

Arnosti, C., Bell, C., Moorhead, D., Sinsabaugh, R., Steen, A., Stromberger, M., *et al.* (2013). Extracellular enzymes in terrestrial, freshwater, and marine environments: perspectives on system variability and common research needs. *Biogeochemistry* **117:** 5-21.

Bauer, M., Kube, M., Teeling, H., Richter, M., Lombardot, T., Allers, E. *et al.* (2006). Whole genome analysis of the marine Bacteroidetes *'Gramella forsetii'* reveals adaptations to degradation of polymeric organic matter. *Environ Microbiol* **8:** 2201-2213.

Berlemont, R., and Martiny, A.C. (2015). Genomic potential for polysaccharide deconstruction in Bacteria. *Appl Environ Microbiol* **81:** 1513-1519.

Beier S, Jones CM, Mohit V, Hallin S, Bertilsson S (2010). Global phylogeography of chitinase genes in aquatic metagenomes. *Appl Environ Microbiol* **77:** 1101-1106.

Cabrera, R., Babul, J., and Guixé, V. (2010). Ribokinase family evolution and the role of conserved residues at the active site of the PfkB subfamily representative, Pfk-2 from *Escherichia coli*. *Arch Biochem Biophys* **502:** 23-30.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res* **37:** D233-D238.

Chung, I.K., Beardall, J., Mehta, S., Sahoo, D., and Stojkovic, S. (2011). Using marine macroalgae for carbon sequestration: a critical appraisal. *J Appl Phycol* **23:** 877-886.

Cordero, O.X., and Datta, M.S. (2016). Microbial interactions and community assembly at microscales. *Curr Opin Microbiol* **31:** 227-234.

Donati, I., and Paoletti, S. (2009). Material properties of alginates. *Alginates: biology and applications*. Springer. pp 1-53.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R. *et al.* (2013). Pfam: the protein families database. *Nucleic Acids Res* **42:** D222-D230.

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28:** 3150-3152.

Garron, M.-L., and Cygler, M. (2010). Structural and mechanistic classification of uronic acid-containing polysaccharide lyases. *Glycobiology* **20:** 1547-1573.

Gerken, H.G., Donohoe, B., and Knoshaug, E.P. (2013). Enzymatic cell wall degradation of *Chlorella vulgaris* and other microalgae for biofuels production. *Planta* **237:** 239-253.

Guibert, L.M., Loviso, C.L., Borglin, S., Jansson, J., Dionisi, H.M. and Lozada, M. (2016). Diverse bacterial groups contribute to the alkane degradation potential of chronically-polluted Subantarctic coastal sediments. *Microb Ecol* **71**:100-112.

Groudieva, T., Grote, R., and Antranikian, G. (2003). *Psychromonas arctica* sp. nov., a novel psychrotolerant, biofilm-forming bacterium isolated from Spitzbergen. *Int J Syst Evol Microbiol* **53:** 539-545.

Hanson, C.A., Fuhrman, J.A., Horner-Devine, M.C., and Martiny, J.B. (2012). Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol* **10:** 497-506.

Hardison, A.K., Canuel, E.A., Anderson, I.C., and Veuger, B. (2010). Fate of macroalgae in benthic systems: carbon and nitrogen cycling within the microbial community. Mar Ecol Prog Ser **414:** 41-55.

Hehemann, J.-H., Boraston, A.B., and Czjzek, M. (2014). A sweet new wave: structures and mechanisms of enzymes that digest polysaccharides from marine algae. *Curr Opin Struct Biol* **28:** 77-86.

Herlemann, D.P.R., Labrenz, M., Jurgens, K., Bertilsson, S., Waniek, J.J., and Andersson, A.F. (2011). Transitions in bacterial communities along the 2000-km salinity gradient of the Baltic Sea. *ISME J* **5:** 1571-1579.

Huntemann, M., Ivanova, N.N., Mavromatis, K., Tripp, H.J., Paez-Espino, D., Tennessen, K. *et al*. (2016). The standard operating procedure of the DOE-JGI Metagenome Annotation Pipeline (MAP v.4). *Stand Genomic Sci* **11:** 17.

Inoue, A., Takadono, K., Nishiyama, R., Tajima, K., Kobayashi, T., and Ojima, T. (2014). Characterization of an alginate lyase, Flalya, from *Flavobacterium* sp. strain UMI-01 and its expression in *Escherichia coli*. *Mar Drugs* **12:** 4693-4712.

Johannesson, K., and André, C. (2006). Life on the margin: genetic isolation and diversity loss in a peripheral marine ecosystem, the Baltic Sea. *Molec Ecol* **15:** 2013-2029.

Kabisch, A., Otto, A., König, S., Becher, D., Albrecht, D., Schüler, M. *et al.* (2014). Functional characterization of polysaccharide utilization loci in the marine Bacteroidetes *Gramella forsetii* KT0803. *ISME J* **8:** 1492-1502.

Kita, A., Miura, T., Kawata, S., Yamaguchi, T., Okamura, Y., Aki, T. *et al.* (2016). Bacterial community structure and predicted alginate metabolic pathway in an alginate-degrading bacterial consortium. *J Biosci Bioeng.* **121:** 286–292.

Kloareg, B., and Quatrano, R. (1988). Structure of the cell walls of marine algae and ecophysiological functions of the matrix polysaccharides. *Oceanography and Marine Biology: an Annual Review.* **26:** 259-315.

Krause-Jensen, D., Duarte, C.M. (2014). Expansion of vegetated coastal ecosystems in the future Arctic. *Front Mar Sci* **1:** 77.

Kunin, V., He, S., Warnecke, F., Peterson, S.B., Martin, H.G., Haynes, M. *et al.* (2008). A bacterial metapopulation adapts locally to phage predation despite global dispersal. *Genome Res* **18:** 293-297.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H. *et al.* (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23:** 2947-2948.

Lee, S.H. (2014). Biochemical and structural characterization of a novel enzyme involved in uronic acid metabolism. Department of Biochemistry and Microbiology, Master of Science. University of Victoria: Victoria, Australia.

Liuzzi, M.G, Gappa, J.L., and Piriz, M.L. (2011). Latitudinal gradients in macroalgal biodiversity in the Southwest Atlantic between 36 and 55°S. *Hydrobiologia* **673:** 205-214.

Lombard, V., Ramulu, H.G., Drula, E., Coutinho, P.M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42:** D490-D495.

Lozada, M., and Dionisi, H.M. (2015). Microbial Bioprospecting in Marine Environments. In: Kim S-K (ed). *Hb25_Springer Handbook of Marine Biotechnology*. Springer. pp 307-326.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

Mackelprang, R., Waldrop, M.P., DeAngelis, K.M., David, M.M., Chavarria, K.L., Blazewicz, S.J. *et al.* (2011). Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. *Nature* **480:** 368-371.

Mann, A.J., Hahnke, R.L., Huang, S., Werner, J., Xing, P., Barbeyron, T. *et al.* (2013). The genome of the alga-associated marine flavobacterium *Formosa agariphila* KMM 3901T reveals a broad potential for degradation of algal polysaccharides. *Appl Environ Microbiol* **79:** 6813-6822.

Markowitz, V.M., Chen, I.-M.A., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y. *et al.* (2012). IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res* **40:** D115-D122.

Markowitz, V.M., Chen, I.-M.A., Chu, K., Szeto, E., Palaniappan, K., Pillay, M. *et al.* (2014). IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucleic Acids Res* **42**: D568-D573.

Mitulla, M., Dinasquet, J., Guillemette, R., Simon, M., Azam, F., and Wietz, M. (2016). Response of bacterial communities from California coastal waters to alginate particles and an alginolytic Alteromonas macleodii strain. *Environ Microbiol*, in press. DOI: 10.1111/1462-2920.13314.

Neumann, A.M., Balmonte, J.P., Berger, M., Giebel, H.-A., Arnosti, C., Voget, S. *et al.* (2015). Different utilization of alginate and other algal polysaccharides by marine *Alteromonas macleodii* ecotypes. *Environ Microbiol* **17**: 3857-3868.

Nogi, Y., Kato, C., and Horikoshi, K. (2002). *Psychromonas kaikoae* sp. nov., a novel from the deepest piezophilic bacterium cold-seep sediments in the Japan Trench. *Int J Syst Evolut Microbiol* **52:** 1527-1532.

Patil, K.R., Roune, L., and McHardy, A.C. (2012). The PhyloPythiaS web server for taxonomic assignment of metagenome sequences. *PLoS ONE* **7:** e38581.

Persson, B., and Kallberg, Y. (2013). Classification and nomenclature of the superfamily of short-chain dehydrogenases/reductases (SDRs). *Chem-Biol Interact* **202:** 111-115.

Quartino, M., and De Zaixso, A.B. (2008). Summer macroalgal biomass in Potter Cove, South Shetland Islands, Antarctica: its production and flux to the ecosystem. *Polar Biol* **31:** 281-294.

Quartino, M.L., de Zaixso, A.B., and Momo, F.R. (2008). Macroalgal production and the energy cycle of Potter Cove. In: Wiencke, C., Ferreyra, G.A., Abele, D. and Marenssi, S. (eds). The Antarctic ecosystem of Potter Cove, King-George Island (Isla 25 de Mayo). Synopsis of research performed 1999-2006 at the Dallmann Laboratory and Jubany Station. pp 68-74.

Quartino, M.L., Deregibus, D., Campana, G.L., Latorre, G.E.J., and Momo, F.R. (2013). Evidence of macroalgal colonization on newly ice-free areas following glacial retreat in Potter Cove (South Shetland Islands), Antarctica. *PLoS ONE* **8:** e58223.

Quartino, M.L., Vazquez, S.C., Latorre, G.E., and Mac Cormack, W.P. (2015). Possible role of bacteria in the degradation of macro algae *Desmarestia anceps* Montagne (Phaeophyceae) in Antarctic marine waters. *Rev Arg Microbiol* **47**: 274–276.

Reddy, V.S., Shlykov, M.A., Castillo, R., Sun, E.I., and Saier, M.H. (2012). The major facilitator superfamily (MFS) revisited. *FEBS J* **279:** 2022-2035.

Rigali, S.B., Derouaux, A., Giannotta, F., and Dusart, J. (2002). Subdivision of the helix-turn-helix GntR family of bacterial regulators in the FadR, HutC, MocR, and YtrA subfamilies. *J Biol Chem* **277:** 12507-12515.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

Wiley-Blackwell and Society for Applied Microbiology

Seon, J., Lee, T., Lee, S.C., Pham, H.D., Woo, H.C., and Song, M. (2014). Bacterial community structure in maximum volatile fatty acids production from alginate in acidogenesis. *Biores Technol* **157**: 22-27.

Sikes, E.L., Uhle, M.E., Nodder, S.D., and Howard, M.E. (2009). Sources of organic matter in a coastal marine environment: evidence from n-alkanes and their ^{13}C distributions in the Hauraki Gulf, New Zealand. *Mar Chem* **113:** 149-163.

Takagi, T., Morisaka, H., Aburaya, S., Tatsukami, Y., Kuroda, K., and Ueda, M. (2016). Putative alginate assimilation process of the marine bacterium *Saccharophagus degradans* 2-40 based on quantitative proteomic analysis. *Mar Biotechnol* **18**: 15-23.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molec Biol Evol* **28:** 2731-2739.

Tang, K., Jiao, N., Liu, K., Zhang, Y., and Li, S. (2012). Distribution and functions of TonB-dependent transporters in marine bacteria and environments: implications for dissolved organic matter utilization. *PLoS ONE* **7:** e41204.

Teske, A., Durbin, A., Ziervogel, K., Cox, C., and Arnosti, C. (2011). Microbial community composition and function in permanently cold seawater and sediments from an Arctic fjord of Svalbard. *Appl Environ Microbiol* **77:** 2008-2018.

Thomas, F., Barbeyron, T., Tonon, T., Génicot, S., Czjzek, M., and Michel, G. (2012). Characterization of the first alginolytic operons in a marine bacterium: from their emergence in marine Flavobacteriia to their independent transfers to marine Proteobacteria and human gut Bacteroides. *Environ Microbiol* **14:** 2379-2394.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

Thureborn P, Lundin D, Plathan J, Poole AM, Sjöberg B-M, Sjöling S (2013). A metagenomics transect into the deepest point of the Baltic Sea reveals clear stratification of microbial functional capacities. *PLoS ONE* **8:** e74983.

Truus, K., Vaher, M., and Taure, I. (2001). Algal biomass from *Fucus vesiculosus* (Phaeophyta): investigation of the mineral and alginate components. *Proc Estonian Acad Sci Chem* **50:** 95-103.

Tuyen, H., Helmke, E., and Schweder, T. (2001). Cloning of two pectate lyase genes from the marine Antarctic bacterium *Pseudoalteromonas haloplanktis* strain ANT/505 and characterization of the enzymes. *Extremophiles* **5:** 35-44.

Vigneron, A., Cruaud, P., Roussel, E.G., Pignet, P., Caprais, J.-C., Callac, N. *et al.* (2014). Phylogenetic and functional diversity of microbial communities associated with subsurface sediments of the Sonora Margin, Guaymas Basin. *PLoS ONE* **9**: e104427.

Wargacki, A.J., Leonard, E., Win, M.N., Regitsky, D.D., Santos, C.N.S., Kim, P.B. *et al.* (2012). An engineered microbial platform for direct biofuel production from brown macroalgae. *Science* **335:** 308-313.

Wiencke, C., and Amsler, C.D. (2012). Seaweeds and their communities in polar regions. *Seaweed Biology*. Springer. pp 265-291.

Wietz, M., Wemheuer, B., Simon, H., Giebel, H.-A., Seibt, M.A., Daniel, R. *et al.* (2015). Bacterial community dynamics during polysaccharide degradation at contrasting sites in the Southern and Atlantic Oceans. *Environ Microbiol* 17: 3822-3831.

Wysocki, L.A., Filley, T.R., and Bianchi, T.S. (2008). Comparison of two methods for the analysis of lignin in marine sediments: CuO oxidation versus tetramethylammonium hydroxide (TMAH) thermochemolysis. *Org Geochem* **39:** 1454-1461.

Xing, P., Hahnke, R.L., Unfried, F., Markert, S., Huang, S., Barbeyron, T. *et al.* (2015). Niches of two polysaccharide-degrading *Polaribacter* isolates from the North Sea during a spring diatom bloom. *ISME J* **9:** 1410-1422.

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., and Xu, Y. (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* **40:** W445-W451.

Zhu, B., and Yin, H. (2015). Alginate lyase: Review of major sources and classification, properties, structure-function analysis and applications. *Bioengineered* **6**: 125-131.

Zinger, L., Boetius, A., and Ramette, A. (2014). Bacterial taxa-area and distance-decay relationships in marine environments. *Molec Ecol* **23:** 954-964.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

Figure Legends

Figure 1. Geographic location of the analyzed sampling regions. NOR, Advent Fjord, Spitsbergen, Svalbard Archipelago, Norway; SWE, Port Värtahamnen, Stockholm, Baltic Sea, Sweden; ARG, Ushuaia Bay, Tierra del Fuego Island, Argentina; ANT, Potter Cove, 25 de Mayo (King George) Island, Antarctica. Two sampling sites distanced ~500 m were selected within each coastal environment, and the top 5 cm of subtidal sediments were sampled in triplicate within each site.

Figure 2. Relative abundance of sequences containing Pfam domains from polysaccharide lyases (PL) in the sediment metagenomes. The estimated copies of sequences containing each of the 12 selected Pfam domains in the unassembled and assembled metagenomes (the latter corrected by read depth) was normalized by dividing by the estimated copies of 12 selected single-copy genes (Table S2). Diameters depict the average value of the 12 calculated ratios. The coefficient of variation across all Pfam domains and samples was <18%. For each PL domain, the Pfam accession number followed by the family name abbreviation between parentheses is indicated. NOR, Svalbard Archipelago (green); SWE, Baltic Sea (orange); ARG, Ushuaia Bay (red); ANT, Potter Cove (blue). Cluster analysis was performed using Bray-Curtis index.

Figure 3. Similarity of metagenomes based on shared alginate lyase homolog (ALH) sequences. Results of ordination analysis (non-metric multidimensional scaling, NMDS) and hierarchical clustering are superimposed in the plot. Samples NOR02, NOR05 and NOR08 (sampling site 1), and samples NOR13, NOR15 and NOR18 (sampling site 2) were retrieved from Advent Fjord, Svalbard Archipelago, Norway. Samples SWE02, SWE07 and SWE12 (KBA site), as well as samples SWE21 and SWE26 (KBB site) were retrieved from the Baltic Sea. Samples ARG01-ARG03 (MC site) and ARG05 (OR site) were retrieved from Ushuaia Bay, Argentina. Samples ARG04 and ARG06 were eliminated from the analysis due to the low

number of identified ALH sequences. ANT01-ANT03 (sampling site 1) and ANT04-ANT06 (sampling site 2), were retrieved from Potter cove, 25 de Mayo (King George) Island, Antarctica (Table S1). Sequences were grouped into operational protein units (OPUs) defined at 80% identity at the amino acid level, and the *sample x OPUs* matrix was subjected to Wisconsin double standardization and used for the construction of the distance matrix, using 1-(Bray-Curtis index) as dissimilarity measure. NMDS stress: 0.09. Circles with filled lines, dotted lines and sparse dotted lines show three grouping levels in clustering analysis, corresponding to 0.8, 0.85 and 0.87 dissimilarity, respectively.

Figure 4. Characteristics of the metagenomic ALH sequence dataset. (A) Evaluation of the level of divergence between ALH sequences and CAZymes. Distribution of E-values obtained with their respective best hits in a blastp analysis, grouped in E-value ranges. (B) Classification of the sequences in CAZy families. *Others:* PL12, PL21, and GH. NC: not classified (too distantly related for a reliable assignment). N: number of identified ALH sequences in each assembled metagenome (with at least 100 amino acids).

Figure 5. Phylogenetic analysis of full-length ALH sequences and scaffold gene organization. Two clusters selected from the phylogenetic tree of ALH sequences from the PL6 family are shown in parts (A) and (B) of the figure (the complete tree is available as Fig. S3). For the metagenomic sequences, the name of the scaffolds are shown in distinct colors depending on the sampling region (green, Svalbard Archipelago; orange, Baltic Sea; blue, Antarctica), and the last number/s in gray correspond to the gene number within the scaffold. Isolates are indicated in bold fonts when the AL sequence was included in the CAZy database, and in normal fonts for sequences from genomes that are not included in the database. The ALH sequences used to build the phylogenetic tree are shown in red in the scaffolds on the right of the figure. The gray rectangle indicates the PL6 catalytic module used for constructing the tree. The rest of the colors in the sequences of the scaffolds indicate orthologous groups (top COG hit,

shown in Fig. S4), except those shown in light yellow that had no COG assignment. Short names of the sequences are indicated in selected scaffolds: *radC*, DNA repair protein RadC; *rraA*, regulator of RNase E activity RraA; *xyll*, xylose isomerase-like TIM barrel; *pl*, PL; *PDG*, 6-phosphogluconate dehydrogenase; *nramp*, Mn^{2+} and Fe^{2+} transporters of the NRAMP family; *kdgF*, cupin domain; *susC*, outer membrane receptor protein; *susD*, SusD family; *pkd*, PKD-domain containing protein; *hyp*, hypothetical protein; *kdgK,* 2-dehydro-3-deoxygluconokinase; *gntR*, GntR family transcriptional regulator; *sdr*, short-chain dehydrogenase/reductase superfamily; *mfs*, major facilitator superfamily transporter; *siae*, sialate O-acetylesterase; *nc*, non-classified ALH; *gh2*, glycoside hydrolase of the GH2 family; tp, long-chain fatty acid transport protein; *G8*, G8-domain containing protein; *kdgA*, 2-keto-3-deoxy-phosphogluconate aldolase; *WD40*, WD40 domain protein beta propeller; *senC*, electron transport protein/SenC; *hk*, histidine kinase; *cgs*, cystathionine gamma-synthase.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

Wiley-Blackwell and Society for Applied Microbiology

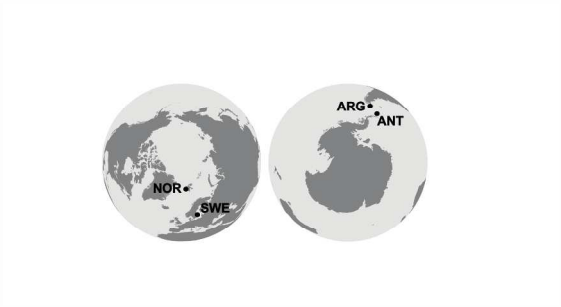

Figure 1. Geographic location of the analyzed sampling regions. NOR, Advent Fjord, Spitsbergen, Svalbard Archipelago, Norway; SWE, Port Värtahamnen, Stockholm, Baltic Sea, Sweden; ARG, Ushuaia Bay, Tierra del Fuego Island, Argentina; ANT, Potter Cove, 25 de Mayo (King George) Island, Antarctica. Two sampling sites distanced ~500 m were selected within each coastal environment, and the top 5 cm of subtidal sediments were sampled in triplicate within each site.
297x420mm (300 x 300 DPI)

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

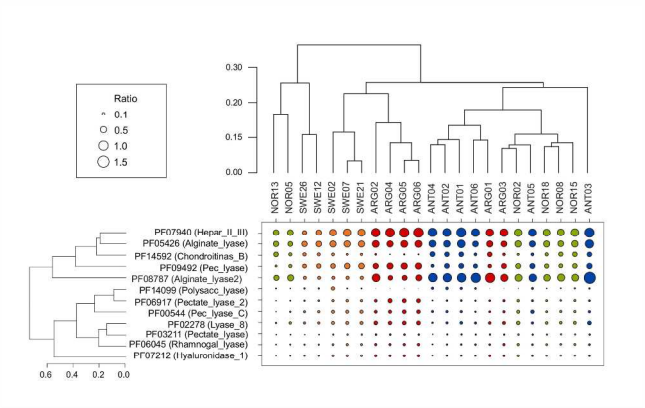

Figure 2. Relative abundance of sequences containing Pfam domains from polysaccharide lyases (PL) in the sediment metagenomes. The estimated copies of sequences containing each of the 12 selected Pfam domains in the unassembled and assembled metagenomes (the latter corrected by read depth) was normalized by dividing by the estimated copies of 12 selected single-copy genes (Table S2). Diameters depict the average value of the 12 calculated ratios. The coefficient of variation across all Pfam domains and samples was <18%. For each PL domain, the Pfam accession number followed by the family name abbreviation between parentheses is indicated. NOR, Svalbard Archipelago (green); SWE, Baltic Sea (orange); ARG, Ushuaia Bay (red); ANT, Potter Cove (blue). Cluster analysis was performed using Bray-Curtis index.
297x420mm (300 x 300 DPI)

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

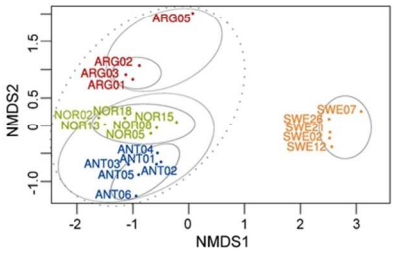

Figure 3. Similarity of metagenomes based on shared alginate lyase homolog (ALH) sequences. Results of ordination analysis (non-metric multidimensional scaling, NMDS) and hierarchical clustering are superimposed in the plot. Samples NOR02, NOR05 and NOR08 (sampling site 1), and samples NOR13, NOR15 and NOR18 (sampling site 2) were retrieved from Advent Fjord, Svalbard Archipelago, Norway. Samples SWE02, SWE07 and SWE12 (KBA site), as well as samples SWE21 and SWE26 (KBB site) were retrieved from the Baltic Sea. Samples ARG01-ARG03 (MC site) and ARG05 (OR site) were retrieved from Ushuaia Bay, Argentina. Samples ARG04 and ARG06 were eliminated from the analysis due to the low number of identified ALH sequences. ANT01-ANT03 (sampling site 1) and ANT04-ANT06 (sampling site 2), were retrieved from Potter cove, 25 de Mayo (King George) Island, Antarctica (Table S1). Sequences were grouped into operational protein units (OPUs) defined at 80% identity at the amino acid level, and the sample x OPUs matrix was subjected to Wisconsin double standardization and used for the construction of the distance matrix, using 1-(Bray-Curtis index) as dissimilarity measure. NMDS stress: 0.09. Circles with filled lines, dotted lines and sparse dotted lines show three grouping levels in clustering analysis, corresponding to 0.8, 0.85 and 0.87 dissimilarity, respectively.
88x37mm (300 x 300 DPI)

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

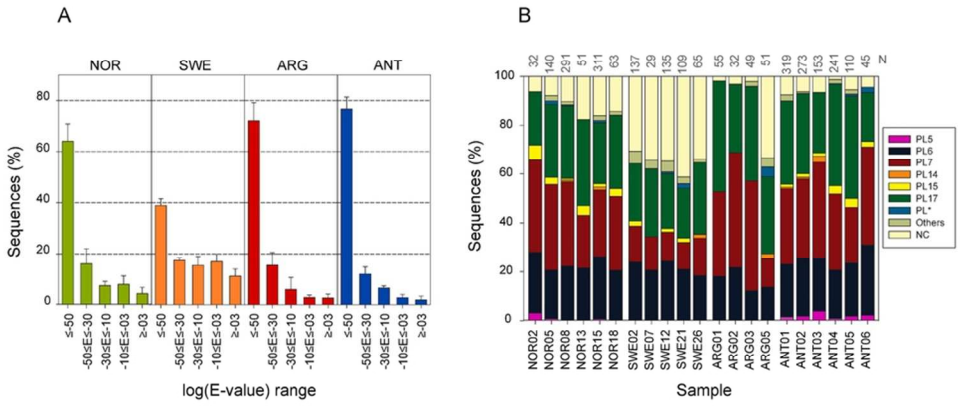

Figure 4. Characteristics of the metagenomic ALH sequence dataset. (A) Evaluation of the level of divergence between ALH sequences and CAZymes. Distribution of E-values obtained with their respective best hits in a blastp analysis, grouped in E-value ranges. (B) Classification of the sequences in CAZy families. Others: PL12, PL21, and GH. NC: not classified (too distantly related for a reliable assignment). N: number of identified ALH sequences in each assembled metagenome (with at least 100 amino acids).
85x41mm (300 x 300 DPI)

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

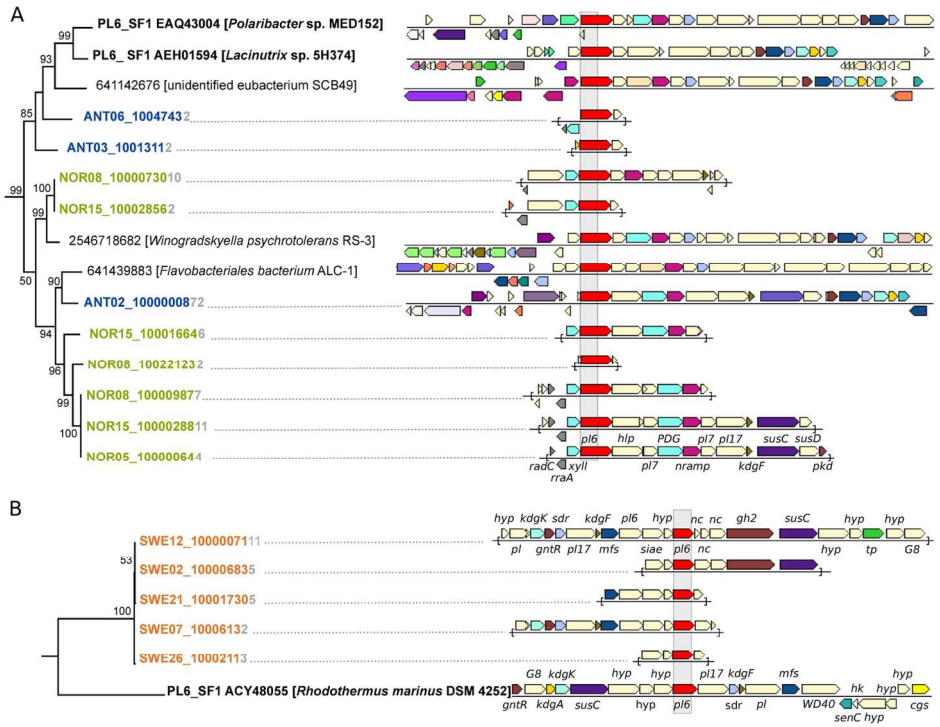

Figure 5. Phylogenetic analysis of full-length ALH sequences and scaffold gene organization. Two clusters selected from the phylogenetic tree of ALH sequences from the PL6 family are shown in parts (A) and (B) of the figure (the complete tree is available as Fig. S3). For the metagenomic sequences, the name of the scaffolds are shown in distinct colors depending on the sampling region (green, Svalbard Archipelago; orange, Baltic Sea; blue, Antarctica), and the last number/s in gray correspond to the gene number within the scaffold. Isolates are indicated in bold fonts when the AL sequence was included in the CAZy database, and in normal fonts for sequences from genomes that are not included in the database. The ALH sequences used to build the phylogenetic tree are shown in red in the scaffolds on the right of the figure. The gray rectangle indicates the PL6 catalytic module used for constructing the tree. The rest of the colors in the sequences of the scaffolds indicate orthologous groups (top COG hit, shown in Fig. S4), except those shown in light yellow that had no COG assignment. Short names of the sequences are indicated in selected scaffolds: radC, DNA repair protein RadC; rraA, regulator of RNase E activity RraA; xyll, xylose isomerase-like TIM barrel; pl, PL; PDG, 6-phosphogluconate dehydrogenase; nramp, Mn2+ and Fe2+ transporters of the NRAMP family; kdgF, cupin domain; susC, outer membrane receptor protein; susD, SusD family; pkd, PKD-domain containing protein; hyp, hypothetical protein; kdgK, 2-dehydro-3-deoxygluconokinase; gntR, GntR family transcriptional regulator; sdr, short-chain dehydrogenase/reductase superfamily; mfs, major facilitator superfamily transporter; siae, sialate O-acetylesterase; nc, non-classified ALH; gh2, glycoside hydrolase of the GH2 family; tp, long-chain fatty acid transport protein; G8, G8-domain containing protein; kdgA, 2-keto-3-deoxy-phosphogluconate aldolase; WD40, WD40 domain protein beta propeller; senC, electron transport protein/SenC; hk, histidine kinase; cgs, cystathionine gamma-synthase.
138x108mm (300 x 300 DPI)

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

Metagenomics unveils the attributes of the alginolytic guilds of sediments from four distant cold coastal environments

Marina N Matos[1], Mariana Lozada[1], Luciano E Anselmino[1], Matías A Musumeci[1], Bernard Henrissat[2,3,4], Janet K Jansson[5], Walter P Mac Cormack[6,7], JoLynn Carroll[8,9], Sara Sjöling[10], Leif Lundgren[11] and Hebe M Dionisi[1]

[1]Laboratorio de Microbiología Ambiental, Centro para el Estudio de Sistemas Marinos (CESIMAR, CONICET), U9120ACD Puerto Madryn, Chubut, Argentina
[2]Architecture et Fonction des Macromolécules Biologiques, CNRS, Aix-Marseille Université, 13288 Marseille, France
[3]INRA, USC 1408 AFMB, F-13288 Marseille, France
[4]Department of Biological Sciences, King Abdulaziz University, Jeddah, 21589, Saudi Arabia
[5]Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA 99352, USA
[6]Instituto Antártico Argentino, Ciudad Autónoma de Buenos Aires, C1064ABR, Argentina
[7]Instituto Nanobiotec, CONICET- Universidad de Buenos Aires, Ciudad Autónoma de Buenos Aires, C1113AAC, Argentina
[8]Akvaplan-niva, Fram – High North Research Centre for Climate and the Environment, NO-9296 Tromsø, Norway
[9]CAGE - Centre for Arctic Gas Hydrate, Environment and Climate, UiT The Arctic University of Norway, N-9037 Tromsø, Norway
[10]School of Natural Sciences and Environmental Studies, Södertörn University, 141 89 Huddinge, Sweden
[11]Stockholm University, SE-106 91 Stockholm, Sweden

SUPPORTING INFORMATION

Table S1. Sampling sites, metadata and metagenome information.

Sampling region	Sampling date	Sampling site	Geographic location	Depth (m)	Temp (ºC)	Salinity (g/l)	Sample	Reads	Assembled reads (%)[1]	CDS unassem.[2]	CDS assem.[3]
Advent Fjord, Spitsbergen, Svalbard Archipelago (NOR)	25-Aug-08	S1	N 78º 14.233' E 15º 40.667'	45	3.75	34.4	NOR02	1.26×10^8	12.99	3.66×10^7	2.82×10^5
							NOR05	2.93×10^8	22.74	4.55×10^7	7.85×10^5
							NOR08	4.47×10^8	29.04	8.25×10^7	1.28×10^6
		S2	N 78º 14.627' E 15º 39.369'	50	3.65	34.5	NOR13	2.48×10^8	17.00	1.21×10^7	3.13×10^5
							NOR15	4.82×10^8	34.55	8.52×10^7	1.47×10^6
							NOR18	2.60×10^8	23.55	4.34×10^7	3.92×10^5
Port Värtahamnen, Baltic Sea, Sweden (SWE)	04-Oct-08	KBA	N 59° 21.800' E 18° 6.690'	21.5	11.20	0.00472	SWE02	3.84×10^8	19.29	9.51×10^7	1.11×10^6
							SWE07	1.34×10^8	8.77	3.23×10^7	2.68×10^5
							SWE12	3.46×10^8	18.17	5.87×10^7	1.08×10^6
		KBB	N 59º 21.667' E 18º 6.291'	21.4	11.40	0.00475	SWE21	2.80×10^8	17.10	7.40×10^7	8.66×10^5
							SWE26	3.44×10^8	12.03	2.74×10^7	5.39×10^5
Ushuaia Bay, Tierra del Fuego, Argentina (ARG)	17-Dec-08	MC	S 54º 48.656' W 68º 17.731'	11.3	8.52	29.42	ARG01	2.00×10^8	4.50	2.37×10^7	1.39×10^5
					8.66	29.34	ARG02	3.07×10^8	5.20	7.67×10^7	1.74×10^5
					8.50	29.42	ARG03	3.19×10^8	13.86	6.95×10^7	4.74×10^5
		OR	S 54º 48.256' W 68º 17.296'	12.3	8.74	29.24	ARG04	1.78×10^8	8.15	4.65×10^7	2.79×10^5
					8.60	29.30	ARG05	3.85×10^8	13.81	9.62×10^7	7.13×10^5
					8.60	29.30	ARG06	1.72×10^8	5.21	4.89×10^7	1.87×10^5
Caleta Potter, 25 de Mayo Island, Antarctica (ANT)	22-Nov-08	S1	S 62º 13.833' W 58º 39.367'	9.5	0.47	34.01	ANT01	3.68×10^8	29.19	7.06×10^7	1.11×10^6
					0.46	34.06	ANT02	3.14×10^8	27.56	5.82×10^7	9.72×10^5
					0.46	34.10	ANT03	1.29×10^8	13.90	3.47×10^7	2.79×10^5
		S2	S 62º 13.917' W 58º 39.300'	23.5	0.19	34.14	ANT04	2.40×10^8	26.45	3.25×10^7	6.57×10^5
					0.20	34.13	ANT05	2.86×10^8	20.40	5.91×10^7	4.67×10^5
					0.20	34.14	ANT06	3.44×10^8	7.80	7.59×10^7	1.01×10^5

[1]Percentage of reads mapping in the scaffolds.
[2]Protein coding sequences in the unassembled fraction of the metagenomes (reads that did not map to scaffolds).
[3]Protein coding sequences in the assembled fraction of the metagenomes.

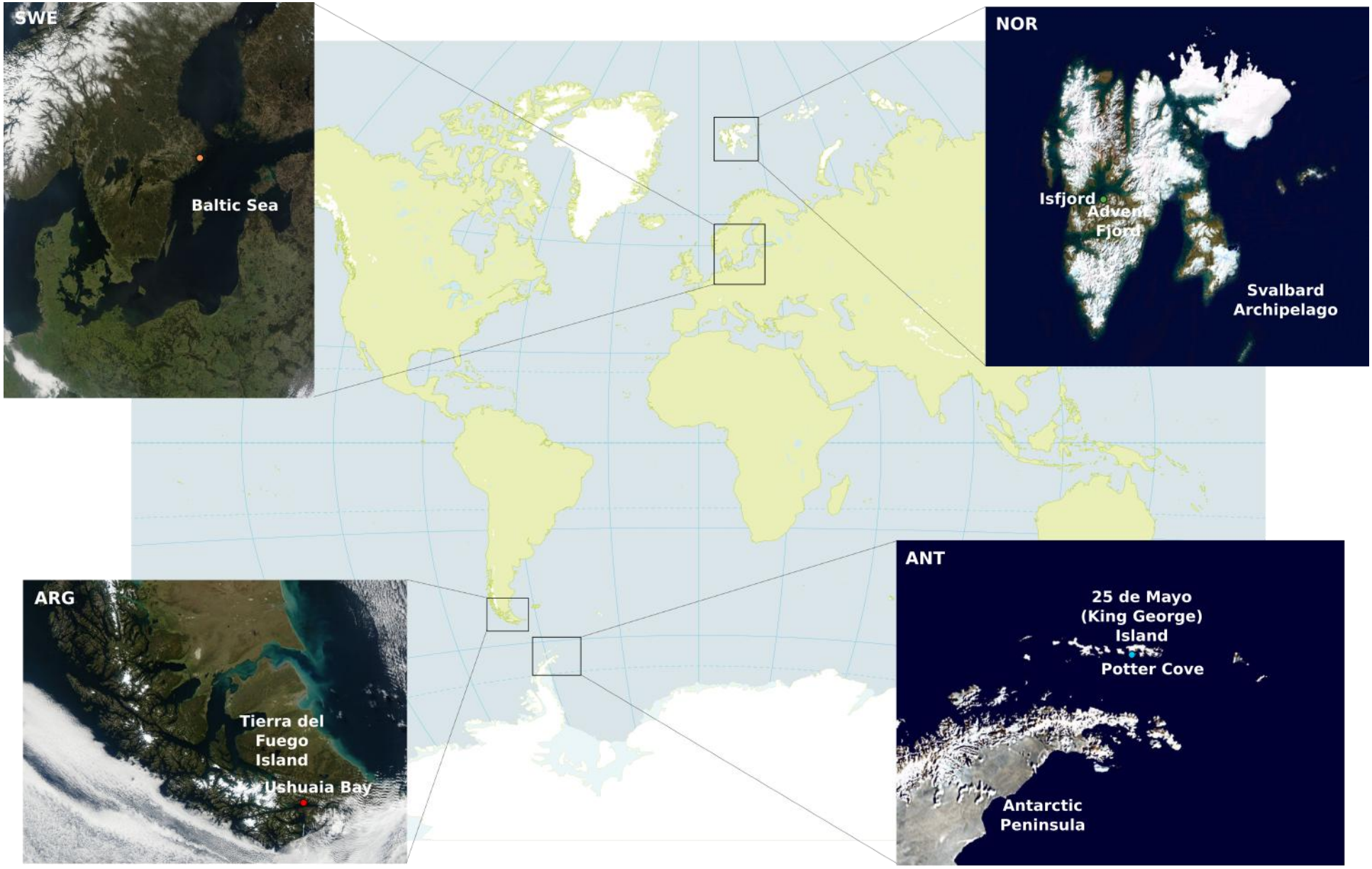

Figure S1. Geographic location of the analyzed sampling regions.

Table S2. List of Pfam domains related to polysaccharide lyases and single-copy ribosomal proteins that were included in the analysis shown in Fig. 2.

Pfam acc. Number	**Abbreviated name [full name]**
PF00544	Pec_lyase_C [Pectate lyase]
PF02278	Lyase_8 [Polysaccharide lyase family 8, super-sandwich domain]
PF03211	Pectate_lyase [Pectate lyase]
PF05426	**Alginate_lyase [Alginate lyase]**
PF06045	Rhamnogal_lyase [Rhamnogalacturonate lyase family]
PF06917	Pectate_lyase_2 [Periplasmic pectate lyase]
PF07212	Hyaluronidase_1 [Hyaluronidase protein (HylP)]
PF07940	**Hepar_II_III [Heparinase II/III-like protein]**
PF08787	**Alginate_lyase2 [Alginate lyase2]**
PF09492	Pec_lyase [Pectic acid lyase]
PF14099	Polysacc_lyase [Polysaccharide lyase]
PF14592	**Chondroitinas_B [Chondroitinase B]**
PF0018	Ribosomal_S3_C [Ribosomal protein, C terminal domain S3]
PF00252	Ribosomal_L16 [Ribosomal protein L16p/L10e]
PF00453	Ribosomal_L20 [Ribosomal protein L20]
PF00542	Ribosomal_L12 [Ribosomal protein L7/L12 C-terminal domain]
PF00831	Ribosomal_L29 [Ribosomal L29 protein]
PF00886	Ribosomal_S16 [Ribosomal protein S16]
PF01016	Ribosomal_L27 [Ribosomal L27 protein]
PF01196	Ribosomal_L17 [Ribosomal protein L17]
PF01245	Ribosomal_L19 [Ribosomal protein L19]
PF01250	Ribosomal_S6 [Ribosomal protein S6]
PF01281	Ribosomal_L9_N [Ribosomal protein L9, N-terminal domain]
PF01649	Ribosomal_S20p [Ribosomal protein S20]

Bold: Pfam domains identified in alginate lyases and oligoalginate lyases.

General features of the study sites

Geological history and current conditions

Sediment samples were collected from four polar or subpolar coastal environments (Fig. S1). Three of the sampling regions are marine environments: (*i*) Advent Fjord (Spitsbergen Island, Svalbard Archipelago, Norway) near the town Longyearbyen, (*ii*) Ushuaia Bay (Tierra del Fuego Island, Argentina) near Ushuaia city, and (*iii*) Potter Cove (25 de Mayo [King George] Island, South Shetland Islands, Antarctica) near the Carlini Antarctic Base. The fourth region is a brackish environment, the Baltic Sea, near Stockholm city. Water-sediment interface temperatures at the time of sampling were 0.2-3.8ºC in the two polar environments (Advent Fjord and Potter Cove), and 8.5-11.4ºC in the two subpolar environments (Baltic Sea and Ushuaia Bay) (Table S1). Although present average water temperatures are currently similar, both polar regions have different geological, oceanographic and climatic histories that must have influenced the evolution of both microbial and seaweed communities (National Research Council, 2003; Wiencke and Amsler, 2012). For instance, the perennial ice cover formed approximately 14 million years ago in Antarctica, while the growth of the Northern Hemisphere ice sheets started 3-5 million years ago (Hansen *et al.*, 2013). In addition, the Antarctic region is highly isolated, while the Arctic is partially connected to nearby temperate regions (Wiencke and Amsler, 2012). Nutrient availability also differs in both regions, being continually high in the Southern Ocean and presenting seasonal variations in the Arctic Ocean. The subpolar marine environment analyzed in this study is Ushuaia Bay, a small inlet located within the Beagle Channel on the south coast of Tierra del Fuego Island (Dionisi *et al.*, 2011). This channel, which connects the Pacific and Atlantic oceans in the south of Argentina and Chile, was covered by ice until approximately 8,000 years ago (Bujalesky, 2007). The second subpolar environment is the Baltic Sea, which is a semi-enclosed and relatively shallow brackish water body connected with the North Sea (Johannesson and Andre, 2006). After a freshwater stage it opened to the North Sea 8,500 years ago, and 4,000 years ago suffered a marine/brackish transition (Pereyra *et al*., 2013). The Baltic Sea is an extreme environment exposed to low winter temperatures, and presents salinity gradients that span more than an order of magnitude (Bergstrom *et al*., 2005; Johannesson and André, 2006; Pereyra *et al*., 2013).

Macroalgae

Polar regions. In spite of the extreme conditions prevalent in polar coastal environments, with constant low temperatures and strong seasonal changes in light conditions, seaweeds thrive in these environments, mostly in sublittoral zones (Wiencke and Amsler, 2012). The Southern Ocean contains many endemic species of macroalgae, while the Arctic Ocean contains fewer endemic species due to its low geographic isolation. Polar environments are characterized by overall low

macroalgae species diversity, although algal diversity is highest around the polar regions selected for this study, Svalbard Archipelago and the Antarctic Peninsula and adjacent islands (Wiencke and Amsler, 2012). In particular, large brown algae species are abundant in these coastal polar environments, where a convergent morpho-functional evolution between members of the order Laminariales in the Arctic and Desmarestiales in Antarctica has been suggested (Wiencke *et al*., 2007). Antarctica is the only region in the world lacking members of the order Laminariales (Wiencke *et al*., 2007). Near the Antarctic Peninsula, *Desmarestia menziesii* typically dominates in shallow waters, while *Desmarestia anceps* dominates in depths up to 30 m and *Himantothallus grandifolius* is more abundant in deeper waters (Quartino and De Zaixso, 2008). The three species represent 80% of the algal biomass in Potter Cove, constituting the main primary producers in this system characterized by a very low phytoplankton biomass accumulation (Quartino and De Zaixso, 2008). Although these species are attached to hard surfaces, an important fraction of their biomass drifts to the soft sediments after detachment, probably supporting a large fraction of the benthos secondary production (Quartino and De Zaixso, 2008). A Southern Hemisphere origin for the family Desmarestiaceae has been suggested, although members of this group are also present in the Arctic Ocean (Peters *et al*., 1997). In fjords of Svalbard, the subtidal communities are dominated by members of the Laminariales order between 5 and 15 m [mainly *Laminaria digitata* and *Saccharina latissima* (formerly *Laminaria saccharina*)], while *Desmarestia aculeata* and *Desmarestia viridis* are frequently found at higher depths (Wulff *et al*., 2009; Tatarek *et al*., 2012). In an study performed in Isfjorden, a larger fjord where Advent Fjord is located, macroalgae detritus has been found to represent an important component of the sedimentary organic matter, significantly contributing to shallow-water and deep-water food webs (Renaud *et al*., 2015).

Ushuaia Bay. On the coast of southern South America, a trend of increased macroalgal diversity with an increase in latitude has been reported, due to the presence of a higher number of Antarctic/sub-Antarctic species in high-latitude environments (Liuzzi *et al.*, 2011). The highest macroalgae diversity was found on the Beagle Channel, where very large forests of the brown algae *Macrocystis pyrifera*, from the order Laminariales, are commonly found (Adami and Gordillo, 1999; Vanella *et al*., 2007). This species, which presents a very high productivity up to 3 kg C m^{-2} yr^{-1} (Velimirov *et al*., 1977), influences the biodiversity and the abundance of other macroalgae in coastal environments (Almanza and Buschmann, 2013). *M. pyrifera* has been found associated with other brown algae, including *Halopteris* sp. and *Ectocarpus* sp., as well as red algae species (Adami and Gordillo, 1999).

Baltic Sea. Due to its complex salinity history, the Baltic Sea contains few species of marine and freshwater macroorganisms, many with specific physiological adaptations (Bergström *et al.*,

2005). The most dominant macroalgal species in the Baltic Sea is *Fucus vesiculosus*, from the order Fucales, with the closely related sympatric species *Fucus radicans* sharing the same ecological environment and depth distribution (Pereyra *et al*., 2013). The abundance of *F. vesiculosus* decreased over the last decades of the last century due to negative effects of anthropogenic activities, although it is currently recovering (Torn *et al.*, 2006; Alexandridis *et al.*, 2012). Most *Fucus* species are highly buoyant due to the presence of air cavities, and they can be transported long distances between sources and sinks after detachment (Rothäusler et al., 2015). Drifting of brown algae detritus is not restricted to members of this genus. For instance, detrital production accounts for more than 80% of kelp productivity (Krumhansl and Scheibling, 2012).

Anthropogenic impacts

The four analyzed coastal environments present some level of anthropogenic impact, such as eutrophication and/or chronic contamination with toxic pollutants such as polycyclic aromatic hydrocarbons (PAHs). Advent Fjord is an Arctic glacier-fed river estuary that that has been impacted by PAHs transported via coal dust from local mining operations. Total PAH concentrations in Advent Fjord sediments are ≤25 μg/g dry weight (dw) (Holte *et al.*, 1996; Evenset *et al*., 2009; Reimann *et al.*, 2009). The other analyzed polar environment, Potter Cove, is exposed to occasional accidental spills and combustion of fuel, mostly related to an Antarctic Base. In spite of year-round activities at this base, the present level of pollution is low, mostly evidenced by PAHs from petrogenic origin in sediments of this fjord-like inlet (≤ 0.21 μg/g dw; Dauner *et al.*, 2015). In addition, both polar environments are experiencing glacial retreat, which is rapidly modifying these ecosystems (Nuth *et al.*, 2013; Quartino *et al.*, 2013).

The Baltic Sea is exposed to multiple anthropogenic impacts including eutrophication and pollution, which has produced an important deterioration of its ecosystem (Korpinen *et al*., 2012). With the exception of Gulf of Bothnia in the north that presents lower deterioration in its environmental conditions, most coastal areas of the Baltic Sea show similar levels of anthropogenic pressures (Korpinen *et al*., 2012; Elmgren *et al.*, 2015). In particular, these impacts are high in the vicinity of large cities such as Stockholm, where the sampling sites were located. Ushuaia city, with a population of near 70,000 people, is located on the coast of Ushuaia Bay. This small bay has an important vessel traffic, a commercial port that receives 900 vessels a year (site MC), an oil jetty used for loading and offloading refined petroleum connected with nearby storage tanks (site OR), and receives the discharge of untreated industrial and domestic wastewater (Dionisi et al., 2011; Commendatore *et al.*, 2012). Chronic pollution with aliphatic hydrocarbons (≤ 1,185 μg/g dw) and PAHs (≤ 0.36 μg/g dw) has been detected in Ushuaia Bay sediments (Commendatore *et al.*, 2012).

Characteristics of the alginate lyase homolog (ALH) sequence dataset

ALH sequences were identified in the assembled metagenomes by blastp searches with representative sequences belonging to PL families that include AL enzymes as queries (see Material and Methods). In addition, the sequences annotated with the product name *alginate lyase* were retrieved. After eliminating sequences with <100 residues and duplicated sequences identified with different approaches or query sequences, 2,705 alginate lyase homolog (ALH) sequences from the 23 assembled metagenomes were selected for further analyses. The number of ALH sequences identified per metagenome varied between 7 (samples ARG04 and ARG06) and 319 (Fig. 4b). A significant correlation (Pearson correlation, $r = 0.88$, $p < 0.001$) was found between the number of ALH sequences retrieved from the metagenomes and the assembly efficiency (number of reads mapping to the scaffolds, Table S1). In addition, the assembly efficiency correlated with the number of reads generated per sample (Pearson correlation, $r = 0.57$, $p = 0.004$). These results suggest that the number of ALH sequences identified per metagenome is affected by a low and variable metagenome coverage. Therefore, all comparative analyses were based on normalized abundances to account for coverage differences.

The ALH sequences identified in the metagenomes from polar sediments (Antarctica and Svalbard) accounted for 75% of the dataset (Fig. S2a). However, when normalized by the number of protein coding sequences in each assembled metagenome, a significantly higher abundance could only be detected in sediments from Antarctica (post-hoc analysis Bonferroni corrected, $p < 0.036$) (Fig. S2b). This result is in agreement with the high relative abundance of Pfam domains associated with AL enzymes detected in the total (assembled and unassembled) metagenomes from Antarctic sediments (Fig. 2).

The majority of the sequences in the dataset were partial, shorter than 200 amino acids. The proportion of short sequences was highest in the metagenomes from Ushuaia Bay sediments (ARG), which presented the lowest assembly efficiency (Fig. S2c and Table S1). In spite of the overall low assembly efficiency, approximately 13% of the full dataset corresponded to complete sequences (Fig. S2a).

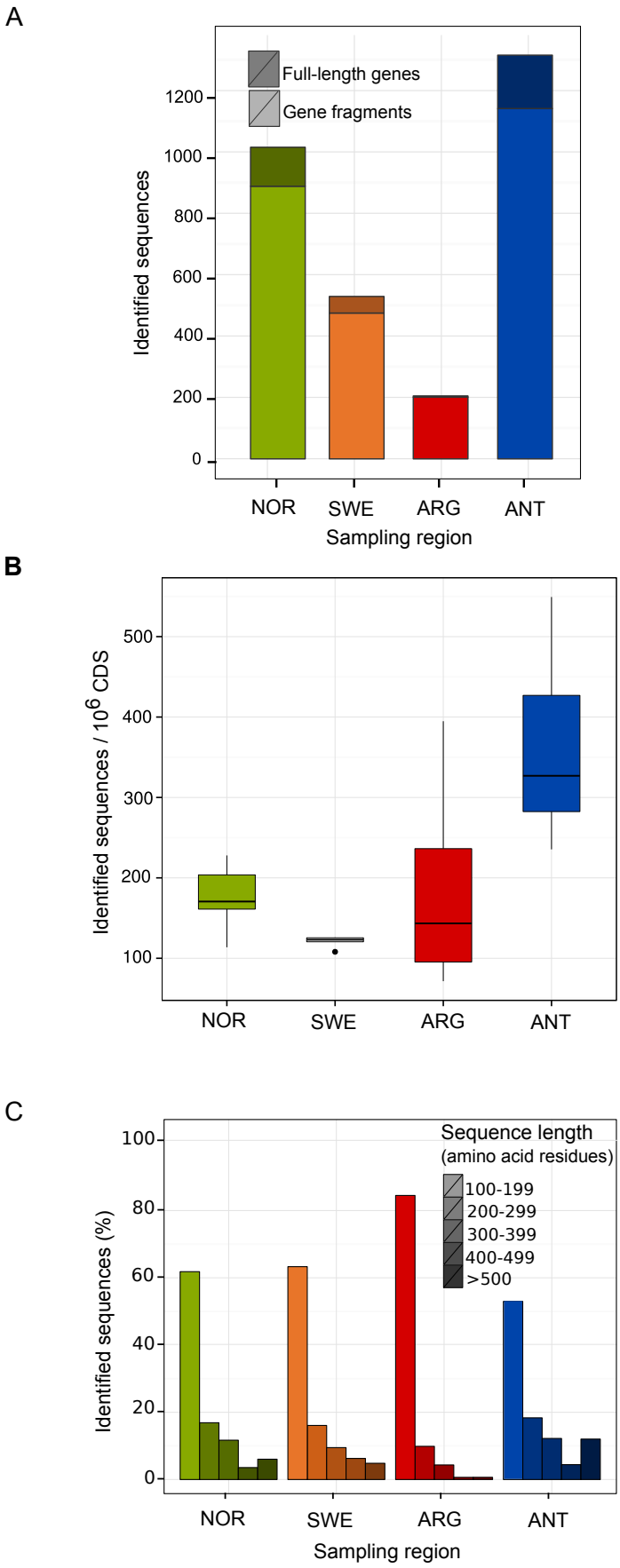

Figure S2. ALH sequences identified per sampling region. (A) Number of identified sequences (full-length and partial) per region. (B) Identified sequences normalized by the number of protein coding sequences (CDS) in each assembled metagenome. The boxplot indicates the median as a line, and whiskers indicate the maximum and minimum observations. (C) Distribution of sequences by length and sampling region, expressed as percentage of total number of sequences identified in each region.

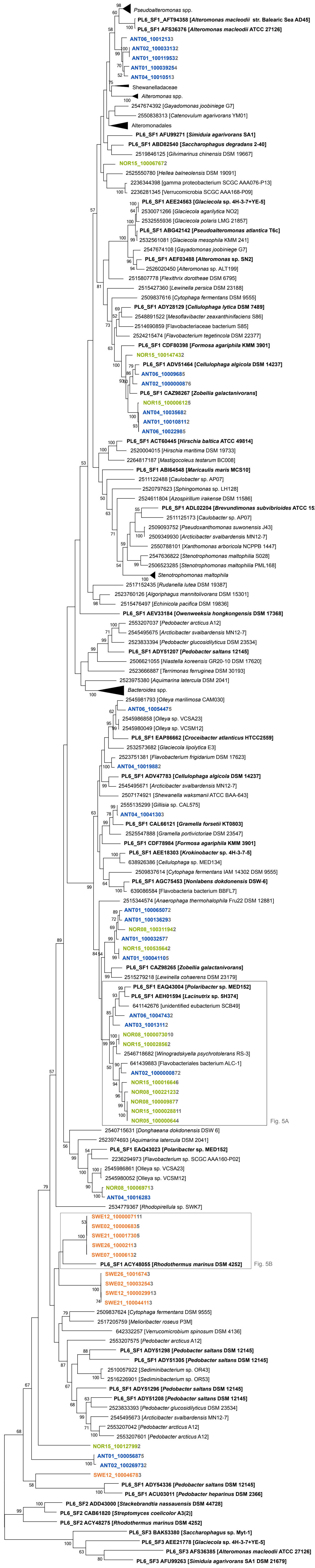

Figure S3. Phylogenetic analysis of PL6 SF1 sequences. The phylogenetic tree (maximum likelihood, WAG amino acid substitution model) was built using the catalytic module of ALH sequences (sequence name in blue, green or orange, for scaffolds from Antarctica, Svalbard and Baltic Sea, respectively; the gene number within the scaffold is shown in gray), AL sequences deposited in the CAZy database (bold font) and homologous sequences from bacterial genomes identified by blastp analysis using CAZymes as query (normal font). The rectangles indicate the clusters selected for the analysis of gene organization of the scaffolds shown in Figure 5. The stability of the tree was evaluated by bootstrap resampling with 500 replications. The bar indicates the inferred

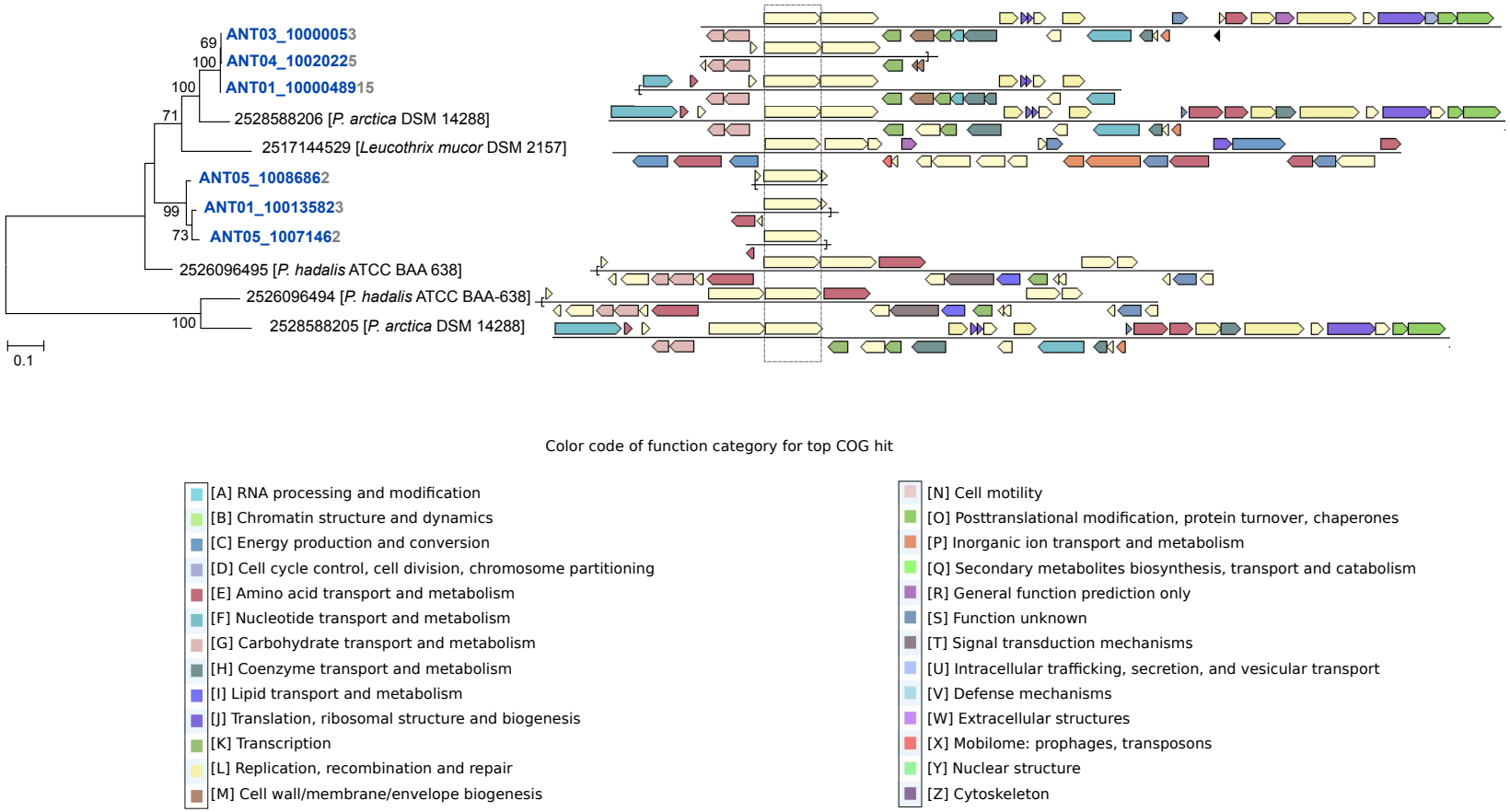

Figure S4. Phylogenetic tree of selected ALH sequences belonging to the PL17 family and related sequences from bacterial genomes. Left, maximum-likelihood tree (LG amino acid substitution model) including selected metagenomic and genomic sequences from the PL17 family, SF2. Bootstrap values were calculated as percentage of 500 replications, and only values > 50% are shown. The scale bar represents the inferred amino acid changes per position. On the right, the scaffolds containing these sequences are depicted. Coloring of the CDS indicate the same orthologous group (top COG hit), except light yellow that denotes no COG assignment. The gene context of the top cluster of ALH sequences from Antarctic metagenomes presents a remarkable conserved synteny with a *P. arctica* genome section containing an homologous gene.

Table S3. Characteristics of unclassified full-length ALH sequences. Only one sequence per cluster (80% identity amino acid level cutoff) was chosen for the analysis.

ALH sequence[1]	Length (aa)	Closest match[2]	Accession Number[2]	Identity (%)[2]	Coverage (%)[2]	Pfam Domain[3]	SP prediction[4]	non-classical secretion[5]
SWE02_100006833[6]	398	hypothetical protein [*Melioribacter roseus* P3M-2]	WP_014856980	73	89	**PF05426**	1-25	
SWE02_100006834[6]	402	hypothetical protein [*Pyrinomonas methylaliphatogenes* K22]	WP_041975089	42	94	**PF05426**	-	-
NOR15_100118034[6]	405	secreted protein [*Asticcacaulis biprosthecium* C19]	WP_006275432	53	84	**PF05426**	-	+
SWE02_100021456	406	hypothetical protein [*Pyrinomonas methylaliphatogenes* K22]	WP_041975089	48	79	**PF05426**	1-26	
SWE02_100036792	382	hypothetical protein [*Flavobacterium frigidarium* DSM 17623]	WP_026706821	65	96	**PF05426**	-	+
SWE02_100183503	201	no hit	-	-	-	no hit	-	-
SWE02_100194032	408	hypothetical protein [*Runella limosa* DSM 17973]	WP_051398284	49	96	**PF05426**	1-22	
SWE02_100429132	262*	alginate lyase [*Niabella* sp. BS26]	ANH82225	52	96	**PF05426**	-	-
SWE02_101983831	149*	alginate lyase [*Draconibacterium orientale* FH5T]	WP_038565039	53	59	**PF05426**	1-40	
SWE12_100000719	234	hypothetical protein [*Pyrinomonas methylaliphatogenes* K22]	WP_041975089	44	94	**PF05426**	-	-

SWE12_100120952	403	probable exported protein YPO3473 [*Jejuia pallidilutea* JCM 19538]	GAL89753	50	94	**PF05426**	1-20	
SWE21_101347302	143	iduronate-2-sulfatase [*Rhodopirellula europaea* 6C]	EMB15970	35	78	no hit	1-18	
ANT01_100105703	304	hypothetical protein [Candidatus *Pelagibacter ubique* HTCC7214]	WP_051624708	33	92	PF14099	1-32	
ANT01_100111993	104	hypothetical protein [*Ruegeria* sp. CECT 5091]	WP_058279887	58	99	no hit	-	-
ANT01_100153122	236	hypothetical protein [*Bdellovibrio bacteriovorus* str. Tiberius]	WP_015091803	35	82	PF14099	-	+
ANT01_100375132	224	hypothetical protein [*Lentisphaera araneosa* HTCC2155]	WP_007279192	55	88	**PF08787**	1-20	
NOR05_101442051	126	hypothetical protein [*Cellulophaga algicola* DSM 14237]	WP_013550178	71	99	no hit	-	-
NOR08_100018094	388	conserved hypothetical protein [*Spirosoma linguale* DSM 74]	ADB36621	49	91	**PF05426**	1-16	
NOR15_100039684	386	alginate lyase [*Siansivirga zeaxanthinifaciens* CC-SAMT-1]	WP_044639167	64	99	**PF05426**	1-16	
NOR15_100056023	409	hypothetical protein [*Marinimicrobium* sp. LS-A18]	WP_024462087	56	96	**PF05426**	1-24	
NOR15_100063937	416	hypothetical protein [*Melioribacter roseus* P3M-2]	WP_014856980	55	93	**PF05426**	1-39	

ANT04_1001209[2]	542	heparinase [*Salinivibrio costicola* ATCC 33508]	WP_021024127	73	99	**PF07940**	-	+
NOR05_100015674	809	heparinase II/III-like protein [*Pyrinomonas methylaliphatogenes* K22]	WP_060635211	27	65	**PF07940**	-	-
NOR15_100327971	689	hypothetical protein AMS26_12205 [*Bacteroides* sp. SM23_62]	KPL14023	49	97	**PF07940**	1-30	

[1]In the sequence number, the scaffold number is indicated in black and the gene number within the scaffold in gray.
[2]Sequence with the highest maximum score in Blastp analysis against the non redundant NCBI database (as for 09-Jun-2016).
[3]Significant Pfam hits (http://pfam.xfam.org); PF05426 (Alginate lyase), PF07940 (Heparinase II/III-like protein), PF08787 (Alginate lyase 2), PF14099 (Polysaccharide lyase).
[4]Prediction of signal peptide (SP) using SingalP 4.1 (http://www.cbs.dtu.dk/services/SignalP/), Phobius (http://phobius.sbc.su.se/), DAS TMfilter (http://mendel.imp.ac.at/sat/DAS/DAS.html) and LipoP 1.0 (http://www.cbs.dtu.dk/services/LipoP/). The residues are indicated when a signal peptide was detected.
[5]Prediction of non-classical protein secretion using SecretomeP 2.0 (http://www.cbs.dtu.dk/services/SecretomeP/).
[6]Sequences with predicted 3-D structure.
*Sequence starts or ends within 3 nucleotides of the start or end of the scaffold.

Three-dimensional models of ALH sequences

Methods. The ALH three-dimensional models were built using the Swiss-Model server (Biasini *et al.*, 2014). The crystal structures of AL enzymes from *Bacteroides ovatus* (PDB 3NNB), *Sphingomonas* sp. A1 (PDB 1HV6) (Yoon *et al.*, 2001) and *Pseudomonas aeruginosa* PAO1 (PDB 4OZV) were used as templates for modeling the structures of sequences NOR15_100118034, SWE02_100064912, SWE02_100006834 and SWE02_100006833. The quality of the models was assessed using QMEAN (Benkert *et al.*, 2008) and VADAR (Willard *et al.*, 2003). Structural superimpositions between templates and models were performed with the software UCSF Chimera (Pettersen *et al.*, 2004) and parameters for the superimposition were calculated using the server PDBeFold of the European Bioinformatics Institute (Mcwilliam *et al.*, 2009). Models with RMSD values of 0.8-1 were considered for further analysis. Structural analyses and image generation were performed using the software UCSF Chimera. The structural stability of the enzymes (both crystallized enzymes and modeled metagenomic sequences) was analyzed using the server PISA (Protein Interfaces, Surfaces and Assemblies) of the European Bioinformatics Institute (Krissinel and Henrick, 2007). The PDB files containing atomic coordinate data for all atoms were used in each case, assuming same cell parameters and space symmetry groups than the corresponding templates for the three dimensional models. Solvent molecules were not included in these analyses.

Results and discussion. We analyzed the three-dimensional relatedness of selected unclassified full-length ALH sequences to AL enzymes with solved structure, in order to evaluate their overall structure and level of conservation in residues known to be critical for AL activity. Two of the modeled metagenomic sequences were SWE02_100064912 and NOR15_100118034, which share 80.1% identity at nucleotide and protein level. The closest matches in the NCBI database were sequences from two proteobacterial strains of the genus *Asticcacaulis* (WP_023448150 and WP_006275432, 47.7 - 50.8% identity at the amino acid level). The lineage of the scaffolds containing the metagenomic sequences suggests that the host of these fragments could also be members of the phylum Proteobacteria (Table S5). The analysis predicted a PL5-type fold similar to the structure of A1-III from *Sphingomonas* sp. A1 (Yoon *et al.*, 2001) for the two sequences, with 12 α-helices arranged as barrel-like structure shaping the substrate binding-site (Fig. S5). AL enzymes that use a histidine/tyrosine mechanism present a highly conserved tyrosine residue in the active site that probably acts as the Brønsted acid in anti β-elimination, or as both acid and base in syn β-elimination (Garron and Cygler, 2010). In an equivalent position, these metagenomic sequences presented a different aromatic residue with moderate hydrophobicity, His251 (Table S4). It is not possible from this analysis to predict the role of this amino acid during the catalysis,

although histidine has been found to act as the Brønsted acid in members of the PL4 family that includes oligogalacturonan lyases (Garron and Cygler, 2014). Another residue that was not conserved in the metagenomic sequences was the histidine that acts as Brønsted base in the β-elimination with an anti configuration (Garron and Cygler, 2010), where a valine residue (Val171) was found instead. Another difference in the predicted structure was the loop formed by the residues 64-85 of A1-III (Yoon *et al.*, 2001), which was determined to play a relevant role as activator of the Brønsted acid Tyr (Mikami *et al.*, 2011). On the other hand, the residue that acts as neutralizer (Asn170) as well as other residues that could participate in substrate interactions (Arg72, Trp123, His250, Tyr254 and Arg309, Table S4) were conserved with respect to A1-III (Yoon *et al.*, 2001). The expression of these sequences will be needed to determine if the differences found in these key residues would affect the catalysis (Mikami *et al.*, 2011), as well as the substrate and the catalytic mechanism of the putative enzymes.

The three-dimensional structures of sequences SWE02_100006833 and SWE02_100006834 were also modeled. These sequences shared only 32.3% identity at the amino acid level, and their predicted structure showed a general fold that was also similar to the one found in members of the PL5 family (Fig. S6). The residues involved in the acid/base mechanism and other relevant amino acids that stabilize the catalytic intermediate were conserved in these sequences (Table S4). However, some differences were found in the amino acids involved in the interaction with the substrate, when compared with characterized AL enzymes of this family (*Azotobacter vinelandii* and *Pseudomonas aeruginosa*; Ertesvåg *et al.*, 1998; MacDonald and Berger, 2014). These residues play an important role in the substrate specificity of the enzyme, because they shape the active site cleft to precisely fit a specific polysaccharide structure (MacDonald and Berger, 2014). Therefore, some differences in the substrate specificity of the modeled alginate lyases with respect to the alginate lyases from this family characterized up till now can be expected.

Interestingly, the four modeled metagenomic sequences presented a lower number of potential hydrogen bonds (SWE02_100064912, 161; NOR15_100118034, 189; SWE02_100006833, 268 and SWE02_100006834, 222) than the three enzymes with known structure of the PL5 family (3NNB, 1HV6 and 4OZV, 365 ± 27.7), from mesophylic microorganisms. This result suggests that these sequences could code for cold-adapted enzymes (Feller and Gerday, 2003).

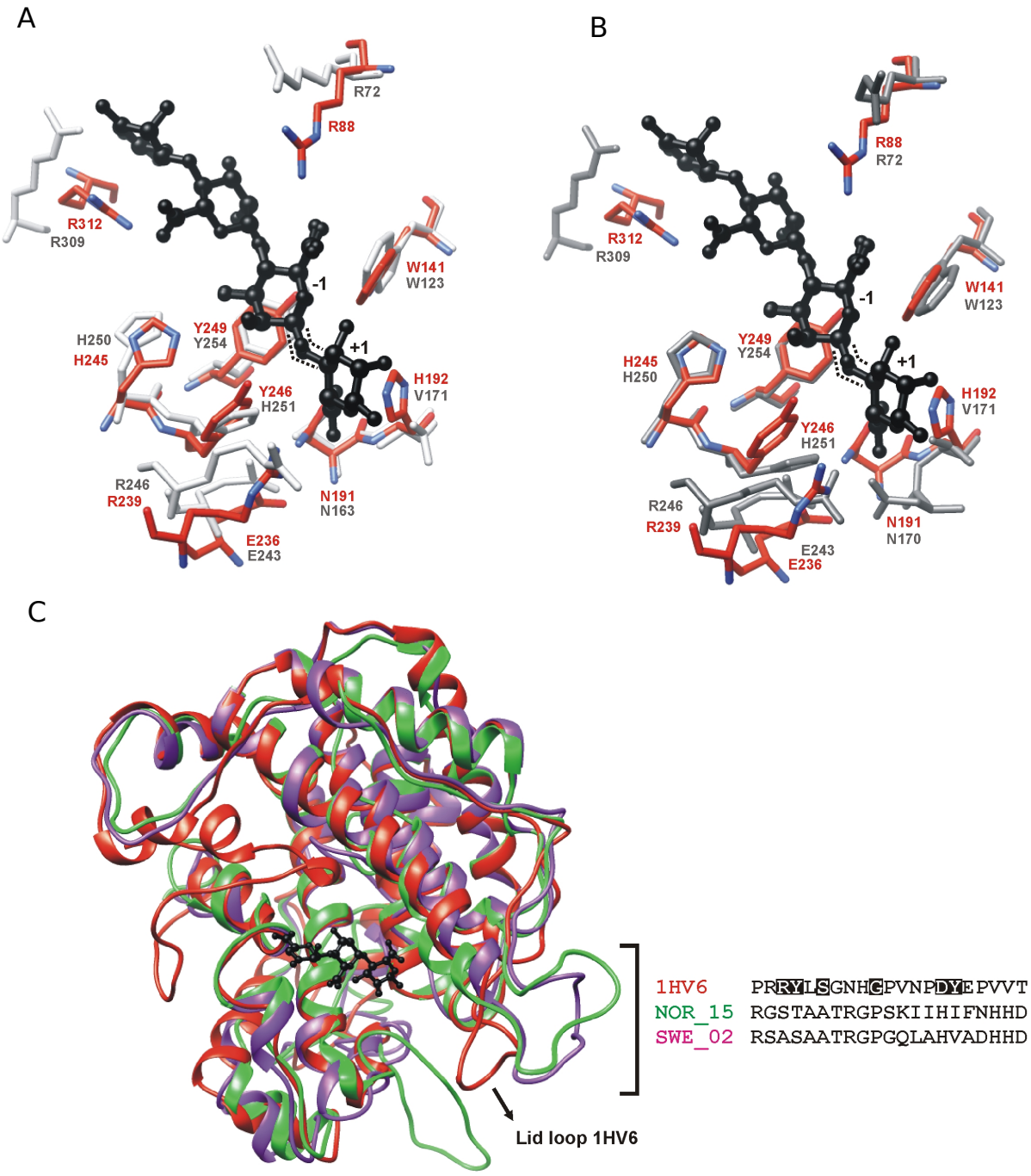

Figure S5. Three-dimensional structural models of the metagenomic sequences NOR15_100118034 and SWE02_00064912. The three-dimensional models of the metagenomic sequences used as template the structure of the alginate lyase BACOVA_01668 from *Bacteroides ovatus* (3NNB). Top: the figure compares the amino acids residues of the active site of alginate lyase A1-III from *Sphingomonas* sp. A1 (1HV6, Yoon *et al*., 2001) with the modeled structures of (A) NOR15_100118034 and (B) SWE02_00064912. An alginate tetrasaccharide substrate (4-deoxy-l-erythro-hex-4-ene-pyranosyluronate-(mannuronate) 2-mannuronic acid) bound to *Sphingomonas* sp. A1 alginate lyase (PDB 4F10, Mikami *et al*., 2011) is shown in black. The bond to be broken between the monosaccharides +1 and -1 is indicated with dashed lines. (C) Structural superimposition of the structure of A1-III from *Sphingomonas* sp. A1 (1HV6, red) with the modeled structures of NOR15_100118034 (green) and SWE02_00064912 (magenta). The trisaccharide product bound to A1-III from *Sphingomonas* sp. A1 is shown in black. To the right, a sequence comparison of the lid loop involved in catalysis is shown. The black boxes correspond to amino acids that are conserved in alginate lyases belonging to the PL5 family.

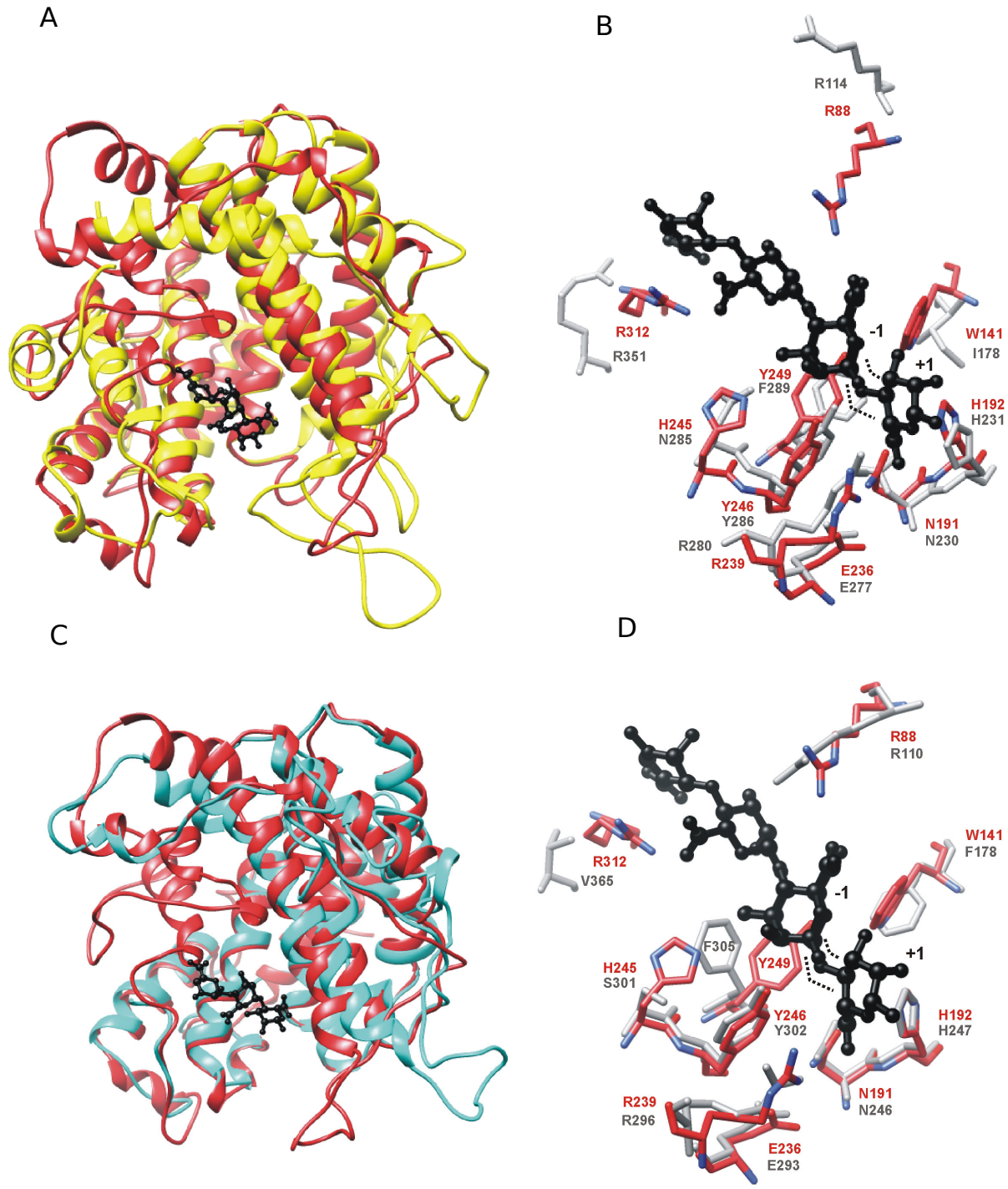

Figure S6. Structural analysis of three-dimensional models of sequences SWE02_100006833 and SWE02_100006834. (A) Structural superposition of the modeled structure of sequence SWE02_100006833 in yellow and alginate lyase A1-III from a *Sphingomonas* sp. A1 in red (1HV6, Yoon *et al*., 2001). The trisaccharide product observed in the crystal structure is shown in black. (B) Main amino acid residues forming the active site of the modeled structure of sequence SWE02_100006833 (grey) and A1-III (red), with a tetrasaccharide substrate represented in black. (C) Comparison of the structural model of SWE02_100006834 (cyan) with A1-III from *Sphingomonas* sp. A1 used as template (red, PDB 4F10, Mikami *et al*., 2011). The alginate tetrasaccharide substrate (4-deoxy-l-erythro-hex-4-ene-pyranosyluronate-(mannuronate) 2-mannuronic acid) bound to the crystal structure of the template is shown in black. (D) Active site of the modeled SWE02_100006834 sequence, compared with its template. Amino acid residues corresponding to templates and modeled structures are identified in red and gray, respectively.

Table S4. Amino acids involved in catalysis[a] and substrate interaction of reported crystal structures of alginate lyases (1HV6, 4OZV and 3NNB) and modeled ALH sequences.

1HV6[b]	**4OZV**[c]	**3NNB**[d]	**NOR15_ 100118034**	**SWE02_ 100064912**	**SWE02_ 100006833**	**SWE02_ 100006834**
Tyr80	Phe85	Leu111	His66	NA	Asn104	NA
Arg88	Thr93	Gly120	Arg72	Arg72	Arg114	Arg110
Gln134	Thr139	Gln168	Lys118	Gln116	Gln161	Asn170
Gln138	Met143	Asp185	Ala121	Leu119	Asp176	Tyr175
Trp141	**Trp146**	**Trp187**	**Trp123**	**Trp123**	Ile178	Phe178
Asn191[a]	**Asn201**[a]	**Asn239**	**Asn170**	**Asn170**	**Asn230**	**Asn246**
His192[a]	**His202**[a]	**His192**	Val171	Val171	**His231**	**His247**
Glu236[a]	**Glu246**[a]	**Glu286**	**Glu243**	**Glu243**	**Glu277**	**Glu293**
Arg239[a]	**Arg249**[a]	**Arg289**	**Arg246**	**Arg246**	**Arg280**	**Arg296**
His245	Ala255	Gly294	His250	His250	Asn285	Ser301
Tyr246[a]	**Tyr256**[a]	**Tyr295**	His251	His251	**Tyr286**	**Tyr302**
Tyr249	Tyr259	Phe298	Tyr254	Tyr254	Phe289	Phe305
Arg306	Thr314	Lys357	Ala307	Thr303	Phe344	Ser359
Arg312	Asn320	Glu359	Arg309	Arg309	Arg351	Val365
Asp314	Lys321	Glu360	Glu314	Phe310	Gln352	Asp364
Arg342	Arg352	NA	NA	NA	NA	NA

[b] 1HV6: crystal structure of alginate lyase from *Sphingomonas sp.* A1 (Yoon *et al.*, 2001)
[c] 4OZV: crystal structure of alginate lyase from *P. aeruginosa* PAO1
[d] 3NNB: crystal structure of alginate lyase from *Bacteroides ovatus*
NA, non-identified amino acid; bold, highly conserved residues.

Table S5. Scaffolds ≥ 4 kb of the sediment metagenomic dataset containing alginate lyase homologs identified in this study.

Scaffold ID[1]	Gene count	Number of ALH sequences	Length (bp)	GC content	Read depth	Lineage (Lineage percentage)[2]	Taxonomic assignment[3]
SA_S1_NOR02_45mDRAFT_c1000894[q]	4	1	4,429	0.31	23	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales (0. 75)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
SA_S1_NOR05_45mDRAFT_c10000064[q]	15	4	20,541	0.33	52	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.53)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Bizionia
SA_S1_NOR05_45mDRAFT_c10001567	5	1	6,871	0.61	92	Unassigned	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Sutterellaceae; Sutterella
SA_S1_NOR05_45mDRAFT_c10003018[t]	7	1	5,268	0.34	30	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales (0.43)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
SA_S1_NOR05_45mDRAFT_c10004408[c]	4	1	4,443	0.57	108	Unassigned	Bacteria; Spirochaetes; Spirochaetia; Spirochaetales; Spirochaetaceae
SA_S1_NOR08_45mDRAFT_c10000173[r]	20	6	23,582	0.33	30	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.7)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
SA_S1_NOR08_45mDRAFT_c10000257[q]	19	4	20,971	0.34	56	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.63)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
SA_S1_NOR08_45mDRAFT_c10000309[s]	17	3	19,635	0.35	59	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales (0.65)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
SA_S1_NOR08_45mDRAFT_c10000583[t]	18	1	15,975	0.37	65	Bacteria; Bacteroidetes (0.67)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Maribacter
SA_S1_NOR08_45mDRAFT_c10000730[u]	13	6	14,715	0.33	30	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales (0.77)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Lacinutrix
SA_S1_NOR08_45mDRAFT_c10000987	13	3	12,893	0.32	59	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales (0.54)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Bizionia
SA_S1_NOR08_45mDRAFT_c10001057[v]	11	1	12,506	0.35	48	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales (0.64)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Maribacter
SA_S1_NOR08_45mDRAFT_c10001316	9	1	11,465	0.33	48	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; unclassified; unclassified; Flavobacteriales bacterium ALC-1 (0.44)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
SA_S1_NOR08_45mDRAFT_c10001563[s]	10	2	10,693	0.36	64	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales (0.8)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae

SA_S1_NOR08_45mDRAFT_c10001809[b]	8	1	10,044	0.52	117	Bacteria (0.75)	Bacteria; Verrucomicrobia; Verrucomicrobiae; Verrucomicrobiales
SA_S1_NOR08_45mDRAFT_c10002309[r]	6	1	9,023	0.33	39	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Cellulophaga (0.5)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Bizionia
SA_S1_NOR08_45mDRAFT_c10003019[w]	8	1	7,977	0.31	28	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales (0.5)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
SA_S1_NOR08_45mDRAFT_c10003370[q]	5	1	7,579	0.34	31	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Cellulophaga (0.6)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
SA_S1_NOR08_45mDRAFT_c10003819[a]	5	1	7,119	0.59	146	Bacteria (0.6)	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae; Jannaschia
SA_S1_NOR08_45mDRAFT_c10004146[r]	4	1	6,841	0.33	23	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales (0.75)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Bizionia
SA_S1_NOR08_45mDRAFT_c10004306	10	1	6,721	0.32	40	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales (0.6)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
SA_S1_NOR08_45mDRAFT_c10004359	5	1	6,684	0.34	39	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales (0.6)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Bizionia
SA_S1_NOR08_45mDRAFT_c10005044[q]	7	2	6,217	0.32	38	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Cellulophaga (0.43)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
SA_S1_NOR08_45mDRAFT_c10006971	5	1	5,286	0.32	27	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales (0.6)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
SA_S1_NOR08_45mDRAFT_c10008236[x]	4	1	4,813	0.39	28	Bacteria (0.75)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Flavobacterium
SA_S1_NOR08_45mDRAFT_c10008444[c]	3	1	4,748	0.56	138	Bacteria (1)	Bacteria; Planctomycetes; Planctomycetia; Planctomycetales; Planctomycetaceae
SA_S1_NOR08_45mDRAFT_c10009369[s]	5	2	4,493	0.34	28	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.6)	Bacteria; Bacteroidetes
SA_S1_NOR08_45mDRAFT_c10010526[y]	4	1	4,209	0.33	55	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.75)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
SA_S1_NOR08_45mDRAFT_c10010672	4	1	4,174	0.59	93	Bacteria (1)	Bacteria

SA_S1_NOR08_45mDRAFT_c10010934	3	1	4,114	0.32	28	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.67)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
SA_S2_NOR15_50mDRAFT_c10000288[q]	15	4	20,029	0.32	46	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales (0.67)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Bizionia
SA_S2_NOR15_50mDRAFT_c10000612	10	2	15,652	0.34	35	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.6)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
SA_S2_NOR15_50mDRAFT_c10000692[y]	12	1	14,930	0.35	60	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales (0.92)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Bizionia
SA_S2_NOR15_50mDRAFT_c10000712[v]	14	1	14,743	0.35	63	Bacteria; Bacteroidetes (0.57)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Maribacter
SA_S2_NOR15_50mDRAFT_c10000915[a]	9	1	13,206	0.58	187	Bacteria (0.67)	Bacteria; Acidobacteria; Candidatus Chloracidobacterium
SA_S2_NOR15_50mDRAFT_c10001309	9	2	11,366	0.34	44	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; unclassified; unclassified; Flavobacteriales bacterium ALC-1 (0.33)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Bizionia
SA_S2_NOR15_50mDRAFT_c10001418[s]	12	2	11,003	0.35	59	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales (0.67)	Bacteria; Bacteroidetes; Cytophagia; Cytophagales; Flammeovirgaceae; Marivirga
SA_S2_NOR15_50mDRAFT_c10001541[b]	8	1	10,620	0.53	105	Bacteria (0.88)	Bacteria; Verrucomicrobia; Verrucomicrobiae; Verrucomicrobiales
SA_S2_NOR15_50mDRAFT_c10001664	7	2	10,280	0.33	35	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.43)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Bizionia
SA_S2_NOR15_50mDRAFT_c10001745[x]	9	1	10,081	0.38	37	Bacteria; Bacteroidetes (0.56)	Bacteria; Bacteroidetes; Flavobacteriia
SA_S2_NOR15_50mDRAFT_c10001786[t]	13	1	10,006	0.37	94	Bacteria; Bacteroidetes; Flavobacteriia (0.54)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Maribacter
SA_S2_NOR15_50mDRAFT_c10002450[q]	12	2	8,751	0.34	39	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.67)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
SA_S2_NOR15_50mDRAFT_c10002856	6	3	8,115	0.33	26	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; unclassified; unclassified; unclassified; Flavobacteriales sp. ALC-1 (0.67)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Lacinutrix
SA_S2_NOR15_50mDRAFT_c10003322[u]	9	4	7,638	0.32	28	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Cellulophaga; Cellulophaga algicola (0.33)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Bizionia
SA_S2_NOR15_50mDRAFT_c10003968	5	1	7,066	0.37	45	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Flavobacterium;	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae

						Flavobacterium johnsoniae (0.8)	
SA_S2_NOR15_50mDRAFT_c10004064	8	1	6,990	0.32	59	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; unclassified; unclassified; unclassified; Flavobacteriales sp. ALC-1 (0.62)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
SA_S2_NOR15_50mDRAFT_c10004344	6	2	6,781	0.5	45	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales (0.67)	Bacteria; Proteobacteria; Gammaproteobacteria; unclassified Gammaproteobacteria; Reinekea
SA_S2_NOR15_50mDRAFT_c10004615[s]	5	2	6,594	0.37	69	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.6)	Bacteria; Bacteroidetes
SA_S2_NOR15_50mDRAFT_c10004658[r]	6	2	6,571	0.34	25	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Kordia; Kordia algicida (0.5)	Bacteria; Proteobacteria; Alphaproteobacteria
SA_S2_NOR15_50mDRAFT_c10005602	6	1	5,993	0.49	37	Bacteria; Proteobacteria (0.67)	Bacteria; Proteobacteria; Gammaproteobacteria
SA_S2_NOR15_50mDRAFT_c10005693[s]	7	2	5,937	0.35	34	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.57)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
SA_S2_NOR15_50mDRAFT_c10005941[r]	4	1	5,815	0.34	26	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Cellulophaga (0.75)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Lacinutrix
SA_S2_NOR15_50mDRAFT_c10006393	8	1	5,627	0.51	46	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales (0.38)	Bacteria; Proteobacteria; Gammaproteobacteria; unclassified Gammaproteobacteria; Reinekea
SA_S2_NOR15_50mDRAFT_c10006767	4	2	5,462	0.47	50	Bacteria; Proteobacteria (0.75)	Bacteria; Proteobacteria; Gammaproteobacteria; unclassified Gammaproteobacteria; Reinekea
SA_S2_NOR15_50mDRAFT_c10006894[w]	3	1	5,409	0.31	23	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; unclassified; unclassified; unclassified; Flavobacteriales sp. ALC-1 (0.67)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
SA_S2_NOR15_50mDRAFT_c10007086	4	1	5,342	0.51	47	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales (0.75)	Bacteria; Proteobacteria; Gammaproteobacteria; unclassified Gammaproteobacteria; Reinekea
SA_S2_NOR15_50mDRAFT_c10007426[r]	5	3	5,228	0.33	24	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; unclassified; unclassified; unclassified; Flavobacteriales sp. ALC-1 (0.6)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Bizionia
SA_S2_NOR15_50mDRAFT_c10007521[r]	6	1	5,196	0.33	25	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.67)	Bacteria; Bacteroidetes; Cytophagia; Cytophagales; Flammeovirgaceae; Marivirga

SA_S2_NOR15_50mDRAFT_c10007930	4	1	5,059	0.34	103	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales (1)	Bacteria
SA_S2_NOR15_50mDRAFT_c10010764[q]	6	1	4,300	0.34	62	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.5)	Bacteria; Proteobacteria
SA_S2_NOR15_50mDRAFT_c10010970[q]	5	2	4,264	0.34	41	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Kordia; algicida; Kordia algicida (0.6)	Bacteria; Bacteroidetes
SA_S2_NOR15_50mDRAFT_c10011148	2	1	4,229	0.49	30	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales (1)	Bacteria; Actinobacteria; Actinobacteria; Micrococcales; unclassified Micrococcineae; Tropheryma
SA_S2_NOR15_50mDRAFT_c10011217[c]	2	1	4,212	0.57	178	Bacteria (1)	Bacteria; Spirochaetes; Spirochaetia; Spirochaetales; Spirochaetaceae; Sphaerochaeta
SA_S2_NOR15_50mDRAFT_c10011803	5	1	4,102	0.52	43	Bacteria (0.8)	Bacteria; Proteobacteria; Gammaproteobacteria; unclassified Gammaproteobacteria; Reinekea
SA_S2_NOR18_50mDRAFT_1000750[a]	5	1	7,650	0,6	212	Bacteria (0.6)	Bacteria; Acidobacteria; Candidatus Chloracidobacterium
SA_S2_NOR18_50mDRAFT_1001546[s]	5	1	5,534	0,34	47	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Cellulophaga; lytica; Cellulophaga lytica (0.6)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
SA_S2_NOR18_50mDRAFT_1001614[c]	3	1	5,449	0,57	223	Bacteria (1)	Bacteria; Planctomycetes; Planctomycetia; Planctomycetales; Planctomycetaceae;m Rhodopirellula
SA_S2_NOR18_50mDRAFT_1002419[s]	6	2	4,475	0,36	36	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.67)	Bacteria; Bacteroidetes
SA_S2_NOR18_50mDRAFT_1002641	5	1	4,269	0,58	139	Unassigned	Bacteria; Actinobacteria; Actinobacteria; Actinomycetales
BS_KBA_SWE02_21mDRAFT_10000262[d]	15	1	18,860	0,36	67	Bacteria (1)	Bacteria; Cloacimonetes; Candidatus Cloacimonas
BS_KBA_SWE02_21mDRAFT_10000683[z]	7	3	12,849	0,35	69	Bacteria; Bacteroidetes (0.57)	Bacteria; Ignavibacteriae; Ignavibacteria; Ignavibacteriales; Ignavibacteriaceae; Ignavibacterium
BS_KBA_SWE02_21mDRAFT_10002018[z]	7	1	7,661	0,34	66	Bacteria; Bacteroidetes (0.71)	Bacteria; Proteobacteria; Alphaproteobacteria; Candidatus Pelagibacter
BS_KBA_SWE02_21mDRAFT_10002145[e]	9	2	7,447	0,33	67	Bacteria; Bacteroidetes (0.44)	Bacteria

BS_KBA_SWE02_21mDRAFT_10002486[g]	7	1	6,879	0,51	37	Bacteria; Proteobacteria; Alphaproteobacteria (0.43)	Bacteria; Proteobacteria; Gammaproteobacteria; unclassified Gammaproteobacteria; Reinekea
BS_KBA_SWE02_21mDRAFT_10003254[d]	6	3	5,959	0,36	54	Bacteria (0.83)	Bacteria; Proteobacteria; Deltaproteobacteria
BS_KBA_SWE02_21mDRAFT_10003679[α]	7	1	5,584	0,32	24	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae;Flavobacterium (0.43)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Flavobacterium
BS_KBA_SWE02_21mDRAFT_10003956[z]	5	1	5,374	0,36	53	Bacteria; Bacteroidetes (0.6)	Bacteria; Ignavibacteriae; Ignavibacteria; Ignavibacteriales; Ignavibacteriaceae; Ignavibacterium
BS_KBA_SWE02_21mDRAFT_10005825	5	1	4,375	0,32	19	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales (0.6)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Aequorivita
BS_KBA_SWE02_21mDRAFT_10006491	6	1	4,116	0,55	22	Bacteria; Proteobacteria; Alphaproteobacteria (0.5)	Bacteria; Proteobacteria; Gammaproteobacteria
BS_KBA_SWE07_21mDRAFT_c1000062[e]	15	1	13,506	0,33	25	Bacteria (0.87)	Bacteria; Proteobacteria; Alphaproteobacteria
BS_KBA_SWE07_21mDRAFT_c1000457[z]	7	2	5,430	0,34	20	Bacteria; Bacteroidetes (0.57)	Bacteria; Ignavibacteriae; Ignavibacteria; Ignavibacteriales; Ignavibacteriaceae; Ignavibacterium
BS_KBA_SWE07_21mDRAFT_c1000613	4	2	4,635	0,35	21	Bacteria (0.75)	Bacteria; Ignavibacteriae; Ignavibacteria; Ignavibacteriales; Ignavibacteriaceae; Ignavibacterium
BS_KBA_SWE12_21mDRAFT_c10000071[z]	23	6	31,457	0,35	56	Bacteria; Bacteroidetes (0.61)	Bacteria; Ignavibacteriae; Ignavibacteria; Ignavibacteriales; Ignavibacteriaceae; Ignavibacterium
BS_KBA_SWE12_21mDRAFT_c10000299[d]	15	3	17,769	0,36	56	Bacteria (0.93)	Bacteria; Cloacimonetes; Candidatus Cloacimonas
BS_KBA_SWE12_21mDRAFT_c10002764[β]	8	1	6,256	0,33	57	Bacteria; Bacteroidetes (0.5)	Bacteria; Firmicutes
BS_KBA_SWE12_21mDRAFT_c10003819[α]	7	1	5,287	0,33	21	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.43)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
BS_KBA_SWE12_21mDRAFT_c10004678	4	1	4,760	0,37	27	Bacteria; Bacteroidetes (1)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
BS_KBA_SWE12_21mDRAFT_c10004840[g]	5	1	4,676	0,5	43	Bacteria; Proteobacteria (0.6)	Bacteria; Proteobacteria; Gammaproteobacteria; unclassified

							Gammaproteobacteria; Reinekea
BS_KBA_SWE12_21mDRAFT_c10005753	4	1	4,271	0,52	32	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales; Alteromonadaceae (0.75)	Bacteria; Proteobacteria; Gammaproteobacteria; unclassified Gammaproteobacteria; Reinekea
BS_KBA_SWE21_205mDRAFT_10000185[d]	15	1	19,618	0.36	46	Bacteria (0.93)	Bacteria; Cloacimonetes; Candidatus Cloacimonas
BS_KBA_SWE21_205mDRAFT_10000614[e]	15	2	11,699	0.33	56	Bacteria (0.8)	Bacteria; Ignavibacteriae; Ignavibacteria; Ignavibacteriales; Ignavibacteriaceae; Ignavibacterium
BS_KBA_SWE21_205mDRAFT_10001442[z]	3	1	7,894	0.36	50	Bacteria; Bacteroidetes (0.67)	Bacteria
BS_KBA_SWE21_205mDRAFT_10001730[z]	6	3	7,246	0.35	49	Bacteria; Bacteroidetes (0.67)	Bacteria; Ignavibacteriae; Ignavibacteria; Ignavibacteriales; Ignavibacteriaceae; Ignavibacterium
BS_KBA_SWE21_205mDRAFT_10002414[α]	7	1	6,135	0.33	18	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.43)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Flavobacterium
BS_KBA_SWE21_205mDRAFT_10003188[g]	6	1	5,277	0.5	35	Bacteria; Proteobacteria (0.67)	Bacteria; Proteobacteria; Gammaproteobacteria; unclassified Gammaproteobacteria; Reinekea
BS_KBA_SWE21_205mDRAFT_10003612[z]	6	1	4,954	0.34	45	Bacteria; Bacteroidetes (0.5)	Bacteria; Proteobacteria; Alphaproteobacteria
BS_KBA_SWE21_205mDRAFT_10004411[d]	4	2	4,450	0.37	41	Bacteria (1)	Bacteria; Proteobacteria
BS_KBB_SWE26_205mDRAFT_c1000167[d]	13	1	17,205	0.37	56	Bacteria; Bacteroidetes (0.54)	Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales
BS_KBB_SWE26_205mDRAFT_c1000211[z]	14	5	14,850	0.34	59	Bacteria; Bacteroidetes (0.57)	Bacteria; Ignavibacteriae; Ignavibacteria; Ignavibacteriales; Ignavibacteriaceae; Ignavibacterium
BS_KBB_SWE26_205mDRAFT_c1000448[z]	5	1	9,334	0.36	48	Bacteria; Bacteroidetes (0.8)	Bacteria; Ignavibacteriae; Ignavibacteria; Ignavibacteriales; Ignavibacteriaceae; Ignavibacterium
BS_KBB_SWE26_205mDRAFT_c1000889[β]	8	1	6,236	0.32	59	Bacteria; Bacteroidetes (0.5)	Bacteria; Firmicutes
BS_KBB_SWE26_205mDRAFT_c1001674[d]	5	2	4,295	0.36	46	Bacteria (0.8)	Bacteria; Proteobacteria

TDF_OR_ARG05_123mDRAFT_1000182	16	1	17,394	0.51	58	Bacteria (0.62)	Bacteria; Proteobacteria; Gammaproteobacteria
KGI_S1_ANT01_95mDRAFT_c10000197[k]	28	1	24,550	0.55	65	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae (0.71)	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae; Octadecabacter
KGI_S1_ANT01_95mDRAFT_c10000389[h]	18	1	19,549	0.57	52	Bacteria; Proteobacteria (0.61)	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae; Celeribacter
KGI_S1_ANT01_95mDRAFT_c10000489[l]	20	2	17,949	0.37	600	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales; Psychromonadaceae; Psychromonas (0.6)	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales
KGI_S1_ANT01_95mDRAFT_c10000719[n]	15	2	15,421	0.35	424	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales; Psychromonadaceae; Psychromonas; ingrahamii; Psychromonas ingrahamii (0.6)	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales
KGI_S1_ANT01_95mDRAFT_c10001023[i]	17	1	13,481	0.58	62	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae (0.53)	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae; Octadecabacter
KGI_S1_ANT01_95mDRAFT_c10001046[ψ]	15	1	13,357	0.34	54	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; unclassified; unclassified; unclassified; Flavobacteriales sp. ALC-1 (0.87)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Bizionia
KGI_S1_ANT01_95mDRAFT_c10001308[μ]	8	1	12,162	0.35	36	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Cellulophaga (0.38)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
KGI_S1_ANT01_95mDRAFT_c10001565	15	1	11,263	0.33	61	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.53)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
KGI_S1_ANT01_95mDRAFT_c10001631[o]	10	1	11,027	0.36	596	Bacteria; Proteobacteria; Gammaproteobacteria (0.7)	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales
KGI_S1_ANT01_95mDRAFT_c10001962[δ]	8	1	10,205	0.35	58	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.38)	Bacteria; Bacteroidetes; Flavobacteriia
KGI_S1_ANT01_95mDRAFT_c10002384[m]	9	2	9,343	0.37	37	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales (0.67)	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales; Colwelliaceae; Colwellia
KGI_S1_ANT01_95mDRAFT_c10002863[φ]	8	3	8,593	0.33	37	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; unclassified; unclassified; Flavobacteriales bacterium ALC-1 (0.5)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Bizionia
KGI_S1_ANT01_95mDRAFT_c10002879	7	2	8,567	0.39	51	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas	Bacteria; Proteobacteria; Gammaproteobacteria; Thiotrichales;

						(0.43)	Piscirickettsiaceae; Cycloclasticus
KGI_S1_ANT01_95mDRAFT_c10003059[λ]	6	2	8,269	0.36	36	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.67)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
KGI_S1_ANT01_95mDRAFT_c10003257[ϖ]	8	1	8,041	0.37	65	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.62)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
KGI_S1_ANT01_95mDRAFT_c10003511[ϕ]	8	2	7,743	0.35	84	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.62)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
KGI_S1_ANT01_95mDRAFT_c10003925[p]	7	2	7,301	0.37	28	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales (0.71)	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales
KGI_S1_ANT01_95mDRAFT_c10003974[μ]	6	1	7,244	0.33	54	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.67)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
KGI_S1_ANT01_95mDRAFT_c10004087	5	1	7,144	0.33	42	Bacteria; Bacteroidetes; Flavobacteriia (0.8)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
KGI_S1_ANT01_95mDRAFT_c10004109	6	2	7,126	0.33	36	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.67)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Flavobacterium
KGI_S1_ANT01_95mDRAFT_c10004110[ϖ]	7	1	7,126	0.38	35	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.57)	Bacteria; Firmicutes; Bacilli; Lactobacillales; Streptococcaceae; Streptococcus
KGI_S1_ANT01_95mDRAFT_c10004141	6	1	7,088	0.35	87	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.67)	Bacteria; Firmicutes; Negativicutes; Selenomonadales; Veillonellaceae
KGI_S1_ANT01_95mDRAFT_c10004161[δ]	6	1	7,077	0.36	43	Bacteria; Bacteroidetes (0.67)	Bacteria; Bacteroidetes
KGI_S1_ANT01_95mDRAFT_c10005039[λ]	4	1	6,403	0.34	70	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Gramella; forsetii; Gramella forsetii (1)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Gillisia
KGI_S1_ANT01_95mDRAFT_c10005318	6	1	6,226	0.38	31	Bacteria (0.67)	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales
KGI_S1_ANT01_95mDRAFT_c10005490[λ]	7	2	6,131	0.34	36	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.57)	Bacteria; Proteobacteria; Alphaproteobacteria; Candidatus Pelagibacter
KGI_S1_ANT01_95mDRAFT_c10005687	6	1	6,017	0.49	32	Bacteria; Verrucomicrobia (0.67)	Bacteria; Proteobacteria; Deltaproteobacteria; Syntrophobacterales; Syntrophaceae; Desulfomonile
KGI_S1_ANT01_95mDRAFT_c10006507[ϖ]	5	1	5,596	0.39	25	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.6)	Bacteria; Bacteroidetes; Flavobacteriia

KGI_S1_ANT01_95mDRAFT_c10007070[γ]	6	2	5,340	0.35	91	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.83)	Bacteria; Bacteroidetes; Cytophagia; Cytophagales; Flammeovirgaceae; Marivirga
KGI_S1_ANT01_95mDRAFT_c10007174	4	1	5,302	0.35	33	Bacteria (0.75)	Bacteria; Bacteroidetes; Flavobacteriia
KGI_S1_ANT01_95mDRAFT_c10007313	8	1	5,236	0.34	63	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.38)	Bacteria; Bacteroidetes; Flavobacteriia
KGI_S1_ANT01_95mDRAFT_c10008112[f]	6	1	4,932	0.36	120	Bacteria (0.83)	Bacteria
KGI_S1_ANT01_95mDRAFT_c10009888[δ]	5	1	4,389	0.35	32	Bacteria; Bacteroidetes (0.6)	Bacteria; Proteobacteria; Alphaproteobacteria
KGI_S1_ANT01_95mDRAFT_c10010444[γ]	5	2	4,250	0.34	36	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.6)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
KGI_S1_ANT01_95mDRAFT_c10010570	5	1	4,219	0.49	33	Bacteria; Proteobacteria; Alphaproteobacteria (0.6)	Bacteria; Proteobacteria; Gammaproteobacteria; Chromatiales; Halothiobacillaceae; Halothiobacillus
KGI_S1_ANT01_95mDRAFT_c10010811	3	2	4,163	0.32	30	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Cellulophaga; algicola; Cellulophaga algicola (0.67)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
KGI_S1_ANT01_95mDRAFT_c10010813	3	1	4,163	0.39	25	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales; Colwelliaceae; Colwellia; psychrerythraea; Colwellia psychrerythraea (0.67)	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales
KGI_S1_ANT01_95mDRAFT_c10011199	7	1	4,086	0.53	55	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae (0.57)	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae
KGI_S1_ANT02_95mDRAFT_c10000008	124	3	136,945	0.33	87	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; unclassified; unclassified; Flavobacteriales bacterium ALC-1 (0.52)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Lacinutrix
KGI_S1_ANT02_95mDRAFT_c10000011	124	1	125,616	0.33	82	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales (0.81)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Lacinutrix
KGI_S1_ANT02_95mDRAFT_c10000038	59	1	62,693	0.33	80	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales (0.76)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Lacinutrix
KGI_S1_ANT02_95mDRAFT_c10000101	38	1	40,410	0.32	78	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales (0.63)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Lacinutrix
KGI_S1_ANT02_95mDRAFT_c10000341[h]	22	1	22,639	0.57	71	Bacteria; Proteobacteria (0.64)	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae; Celeribacter

KGI_S1_ANT02_95mDRAFT_c10000358[ϕ]	17	2	22,229	0.34	81	Bacteria; Bacteroidetes; Flavobacteriia (0.59)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
KGI_S1_ANT02_95mDRAFT_c10000813[l]	17	2	15,086	0.37	211	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales; Psychromonadaceae; Psychromonas (0.59)	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales
KGI_S1_ANT02_95mDRAFT_c10001028[i]	19	1	13,530	0.58	81	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae (0.53)	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae; Octadecabacter
KGI_S1_ANT02_95mDRAFT_c10001460[n]	9	1	11,214	0.36	220	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales; Psychromonadaceae; Psychromonas; ingrahamii; Psychromonas ingrahamii (1)	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales
KGI_S1_ANT02_95mDRAFT_c10001899[φ]	8	3	9,745	0.33	30	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; unclassified; unclassified; unclassified; Flavobacteriales bacterium ALC-1 (0.62)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Bizionia
KGI_S1_ANT02_95mDRAFT_c10002253[ε]	8	2	8,837	0.35	57	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Kordia; Kordia algicida (0.38)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
KGI_S1_ANT02_95mDRAFT_c10002299	7	1	8,736	0.36	30	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas; mendocina; Pseudomonas mendocina (0.71)	Bacteria; Proteobacteria; Gammaproteobacteria
KGI_S1_ANT02_95mDRAFT_c10003313[p]	6	2	7,150	0.34	41	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales (0.5)	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales; Colwelliaceae; Colwellia
KGI_S1_ANT02_95mDRAFT_c10004292[θ]	9	2	6,206	0.33	28	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; unclassified; unclassified; unclassified; Flavobacteriales bacterium ALC-1 (0.56)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Lacinutrix
KGI_S1_ANT02_95mDRAFT_c10004467[j]	8	1	6,071	0.49	39	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales (0.38)	Bacteria; Proteobacteria; Deltaproteobacteria; Desulfuromonadales; Desulfuromonadaceae; Desulfuromonas
KGI_S1_ANT02_95mDRAFT_c10004856[π]	3	1	5,785	0.34	76	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Flavobacterium; johnsoniae; Flavobacterium johnsoniae (0.67)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Flavobacterium
KGI_S1_ANT02_95mDRAFT_c10004857[m]	5	2	5,785	0.38	48	Bacteria; Proteobacteria; Gammaproteobacteria (0.6)	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales
KGI_S1_ANT02_95mDRAFT_c10005348[δ]	4	1	5,475	0.36	26	Bacteria; Bacteroidetes (0.75)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae

KGI_S1_ANT02_95mDRAFT_c10006599[j]	8	1	4,845	0.5	33	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales (0.38)	Bacteria; Proteobacteria; Deltaproteobacteria; Desulfobacterales; Desulfobacteraceae; Desulfobacterium
KGI_S1_ANT02_95mDRAFT_c10006739	5	1	4,788	0.35	21	Bacteria (0.6)	Bacteria; Cloacimonetes
KGI_S1_ANT02_95mDRAFT_c10007123	6	1	4,638	0.34	36	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.5)	Bacteria; Bacteroidetes; Flavobacteriia
KGI_S1_ANT02_95mDRAFT_c10007982[o]	6	1	4,336	0.38	206	Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales; Vibrionaceae; Vibrio (0.67)	Bacteria; Proteobacteria; Gammaproteobacteria
KGI_S1_ANT02_95mDRAFT_c10008111	5	2	4,301	0.35	30	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Gramella; forsetii; Gramella forsetii (0.6)	Bacteria
KGI_S1_ANT02_95mDRAFT_c10008677	3	1	4,129	0.36	41	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales; Pseudoalteromonadaceae; Pseudoalteromonas (0.67)	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales
KGI_S1_ANT03_95mDRAFT_c1000005[l]	37	1	37,200	0.37	66	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales; Psychromonadaceae; Psychromonas (0.57)	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales
KGI_S1_ANT03_95mDRAFT_c1000027[n]	23	1	23,767	0.36	64	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales; Psychromonadaceae; Psychromonas; ingrahamii; Psychromonas ingrahamii 37 (0.65)	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales
KGI_S1_ANT03_95mDRAFT_c1000078[o]	15	1	15,147	0.36	62	Bacteria; Proteobacteria; Gammaproteobacteria (0.67)	Bacteria; Proteobacteria; Gammaproteobacteria
KGI_S1_ANT03_95mDRAFT_c1000421	7	1	6,412	0.31	34	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.71)	Bacteria; Tenericutes; Mollicutes
KGI_S1_ANT03_95mDRAFT_c1000583[λ]	6	2	5,433	0.35	32	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Gramella; Gramella forsetii (0.5)	Bacteria; Firmicutes; Bacilli; Lactobacillales
KGI_S1_ANT03_95mDRAFT_c1000894	4	1	4,327	0.31	27	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales (0.75)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
KGI_S2_ANT04_2345mDRAFT_c1000236	24	2	23,479	0.37	50	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales (0.58)	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales
KGI_S2_ANT04_2345mDRAFT_c1000417[n]	20	1	18,199	0.35	218	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales; Psychromonadaceae; Psychromonas; ingrahamii; Psychromonas ingrahamii (0.7)	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales

KGI_S2_ANT04_2345mDRAFT_c1000939[ε]	13	2	12,453	0.35	68	Bacteria; Bacteroidetes (0.62)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
KGI_S2_ANT04_2345mDRAFT_c1001051[p]	11	2	11,698	0.39	50	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales (0.64)	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales; Colwelliaceae; Colwellia
KGI_S2_ANT04_2345mDRAFT_c1001179[o]	10	1	11,071	0.36	229	Bacteria; Proteobacteria; Gammaproteobacteria (0.8)	Bacteria; Proteobacteria; Gammaproteobacteria
KGI_S2_ANT04_2345mDRAFT_c1001209	9	1	10,927	0.34	31	Bacteria; Proteobacteria; Gammaproteobacteria (0.33)	Bacteria; Proteobacteria; Gammaproteobacteria
KGI_S2_ANT04_2345mDRAFT_c1001273	13	1	10,612	0.4	46	Bacteria (0.62)	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales
KGI_S2_ANT04_2345mDRAFT_c1001398	10	2	10,118	0.35	37	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.9)	Bacteria
KGI_S2_ANT04_2345mDRAFT_c1001628[ψ]	7	1	9,381	0.34	52	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; unclassified; unclassified; unclassified; Flavobacteriales bacterium ALC-1 (0.71)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Bizionia
KGI_S2_ANT04_2345mDRAFT_c1001874[λ]	7	3	8,734	0.36	31	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.71)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
KGI_S2_ANT04_2345mDRAFT_c1001918[ϕ]	9	2	8,612	0.35	67	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.56)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
KGI_S2_ANT04_2345mDRAFT_c1001988[π]	6	1	8,446	0.35	79	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Flavobacterium (0.5)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Flavobacterium
KGI_S2_ANT04_2345mDRAFT_c1002022[l]	9	2	8,369	0.38	221	Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales; Vibrionaceae; Vibrio; Vibrio sinaloensis (0.33)	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales
KGI_S2_ANT04_2345mDRAFT_c1002335	5	2	7,757	0.34	31	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Cellulophaga; algicola; Cellulophaga algicola (0.6)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Lacinutrix
KGI_S2_ANT04_2345mDRAFT_c1003568[ξ]	4	2	5,978	0.33	35	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Cellulophaga; algicola; Cellulophaga algicola (0.75)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
KGI_S2_ANT04_2345mDRAFT_c1003963	3	2	5,592	0.34	34	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.67)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Bizionia
KGI_S2_ANT04_2345mDRAFT_c1004075[γ]	6	2	5,489	0.35	56	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.83)	Bacteria; Bacteroidetes; Cytophagia; Cytophagales; Flammeovirgaceae; Marivirga

KGI_S2_ANT04_2345mDRAFT_c1004130	4	1	5,436	0.33	35	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (1)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Gillisia
KGI_S2_ANT04_2345mDRAFT_c1005207	5	2	4,706	0.33	20	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; unclassified; unclassified; unclassified; Flavobacteriales bacterium ALC-1 (0.6)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Lacinutrix
KGI_S2_ANT04_2345mDRAFT_c1005830	5	2	4,339	0.36	53	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae (0.8)	Bacteria; Bacteroidetes; Flavobacteriia
KGI_S2_ANT04_2345mDRAFT_c1006474	3	1	4,043	0.37	26	Bacteria (1)	Bacteria; Firmicutes
KGI_S2_ANT05_2345mDRAFT_c1000420	18	1	12,867	0.4	101	Bacteria; Proteobacteria (0.56)	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales
KGI_S2_ANT05_2345mDRAFT_c1002760	4	1	5,291	0.44	73	Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales; Vibrionaceae (0.75)	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales; Psychromonadaceae; Psychromonas
KGI_S2_ANT06_2345mDRAFT_c1000624[h]	12	1	13,653	0.55	48	Unassigned	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae; Octadecabacter
KGI_S2_ANT06_2345mDRAFT_c1000968	8	2	10,673	0.32	32	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Cellulophaga; Cellulophaga algicola (0.38)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Lacinutrix
KGI_S2_ANT06_2345mDRAFT_c1001213	5	2	9,535	0.36	27	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales (1)	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales; Colwelliaceae; Colwellia
KGI_S2_ANT06_2345mDRAFT_c1001996	11	1	7,171	0.36	64	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales; Psychromonadaceae; Psychromonas; ingrahamii; Psychromonas ingrahamii (0.55)	Bacteria; Proteobacteria; Gammaproteobacteria
KGI_S2_ANT06_2345mDRAFT_c1002298[ξ]	6	1	6,669	0.33	33	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Cellulophaga; Cellulophaga algicola (0.5)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
KGI_S2_ANT06_2345mDRAFT_c1002488	4	1	6,431	0.44	35	Bacteria; Proteobacteria (0.75)	Bacteria; Proteobacteria; Gammaproteobacteria
KGI_S2_ANT06_2345mDRAFT_c1002697	5	1	6,139	0.49	27	Bacteria; Verrucomicrobia (0.6)	Bacteria; Proteobacteria; Deltaproteobacteria; Syntrophobacterales; Syntrophaceae; Desulfomonile
KGI_S2_ANT06_2345mDRAFT_c1002792	7	1	6,033	0.33	32	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; unclassified; unclassified; Flavobacteriales bacterium ALC-1 (0.57)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae

KGI_S2_ANT06_2345mDRAFT_c1004355[k]	6	1	4,706	0.54	59	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae (0.67)	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae
KGI_S2_ANT06_2345mDRAFT_c1004743	4	2	4,479	0.32	35	Bacteria; Bacteroidetes (1)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Mesoflavibacter
KGI_S2_ANT06_2345mDRAFT_c1005440	2	2	4,140	0.34	32	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; unclassified; unclassified; unclassified; Flavobacteriales bacterium ALC-1 (1)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Bizionia
KGI_S2_ANT06_2345mDRAFT_c1005447[θ]	5	2	4,136	0.33	30	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales (0.6)	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Bizionia

[1]Scaffolds with similar gene organization are indicated with the same superscript letter after the scaffold ID, the scaffolds without a letter are unique within the selected scaffolds (≥4 kb), containing alginate lyase homologs.
[2]Lineage assigned to the scaffolds in the functional annotation [IMG/M pipeline (Markowitz *et al.*, 2014)].
[3]Taxonomic assignment using the composition-based taxonomic classifier PhylopythiaS (Patil *et al.*, 2012), model generic genus 800.

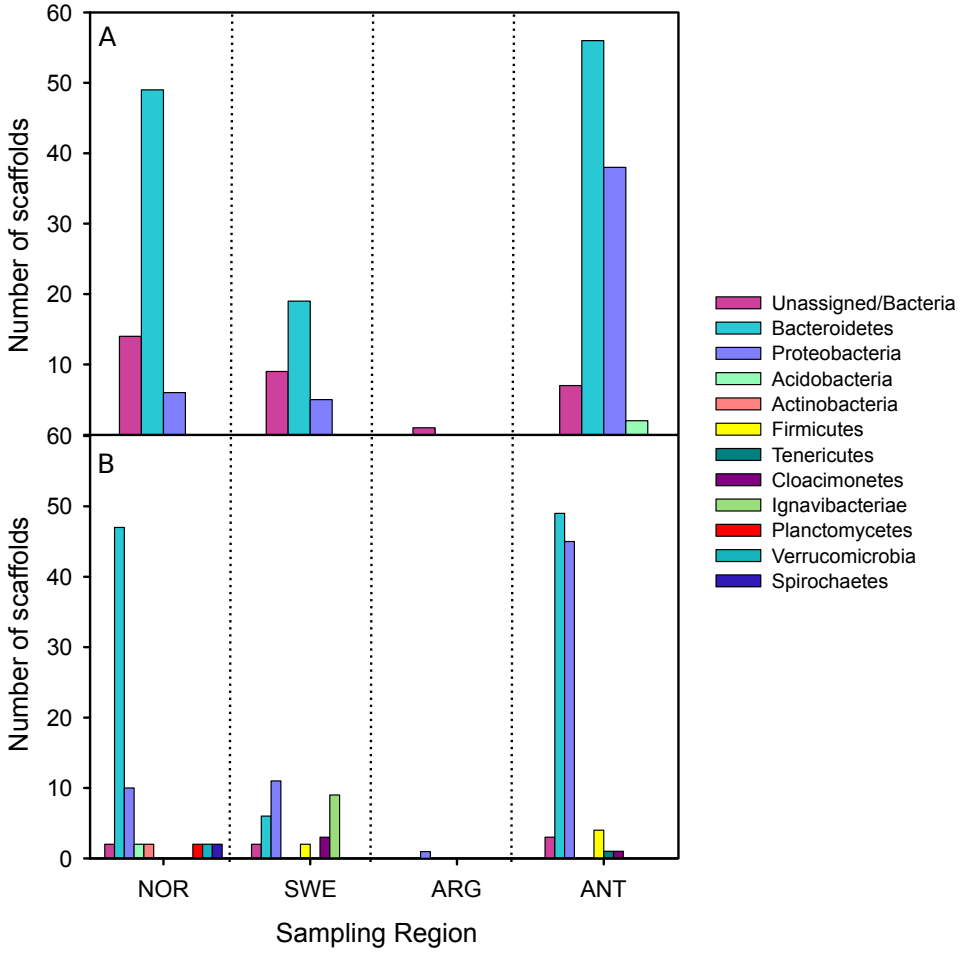

Figure S7. Number of scaffolds ≥4 kb containing ALH sequences, discriminated by sampling region and by lineage. (A) Taxonomic assigment performed in the IMG/M pipeline; (B) Taxonomic assignment using PhylopythiaS. The taxonomic assignment of each scaffold can be found in Supplementary Table S5. NOR: Svalbard Archipelago; SWE: Baltic Sea; ARG: Ushuaia Bay, Argentina; ANT: Potter Cove, Antarctica.

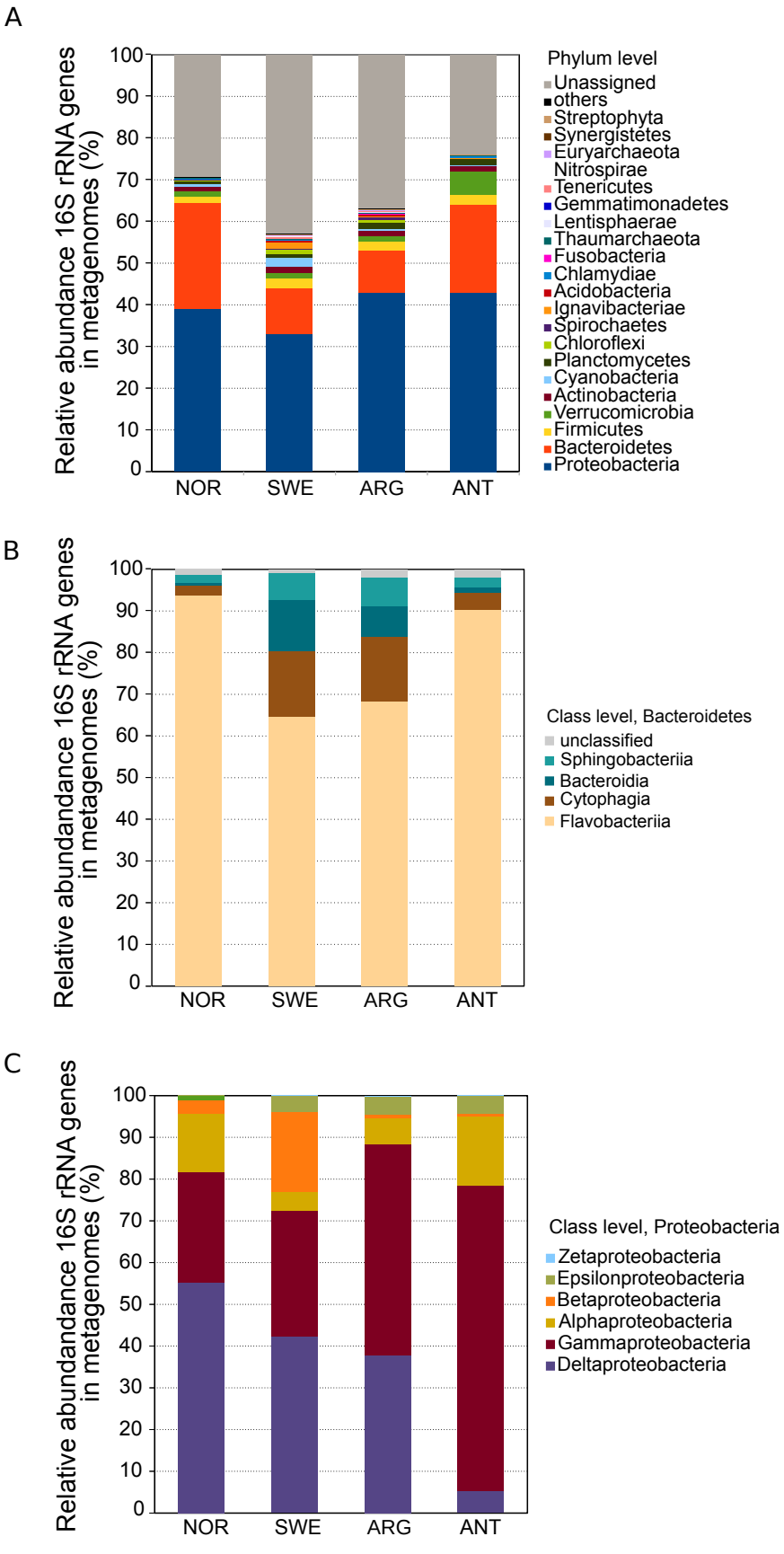

Figure S8. Microbial community structure of cold sediments. The relative abundance of the different taxa are based on blast results (≥90% identity) of 16S rRNA genes in the sediment metagenomes (average values per sampling region, 5-6 samples). (A) Relative abundance at the phylum level (average coefficient of variation 44%). (B) Relative abundance of Bacteroidetes classes (average coefficient of variation 47%). (C) Relative abundance of Proteobacteria classes (average coefficient of variation 38%).

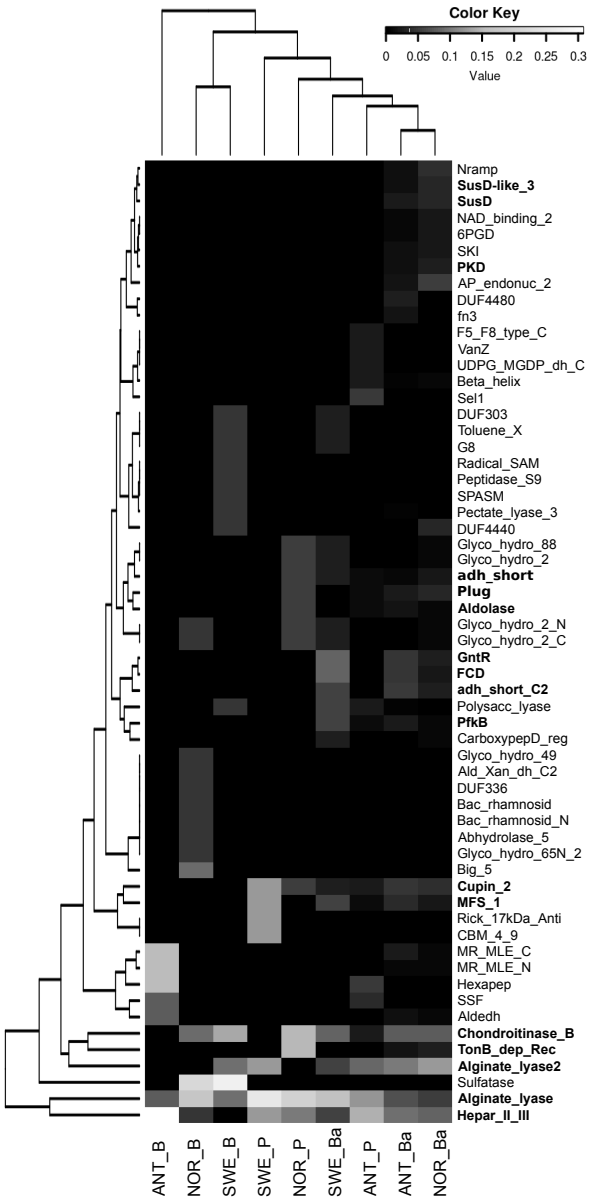

Figure S9. Genomic context of metagenomic ALH sequences in selected scaffolds. The heatmap shows significant Pfam domain hits (gathering threshold) of the sequences most frequently found in the vicinity of ALH sequences. Only the genes codirectional to the identified ALH sequences in scaffolds ≥4 kb were taken into account in the analysis (only one scaffold from ARG samples met these criteria, which was not included in the analysis). The information was discriminated by region and by taxonomic assignment (phyla or superphyla) of the scaffold as assigned in the functional annotation (IMG/M pipeline). The analysis included only the hits that were detected with a frequency > 0.015 in at least one set (taxa and sampling region), and did not take into account sequences without significant Pfam hits (17% of the sequences of the gene clusters). Clustering was performed based on Bray-Curtis similarity index calculated for the gene x sample matrix, previously standardized by Wisconsin double standardization. B: Bacteria and unclassified; Ba: Bacteroidetes; P: Proteobacteria; NOR, Advent Fjord, Svalbard Archipelago, Norway; SWE, Baltic Sea, Sweden; ANT, Potter Cove, 25 de Mayo Island, Antarctica. In bold, Pfam domains of genes associated with alginate utilization processes.

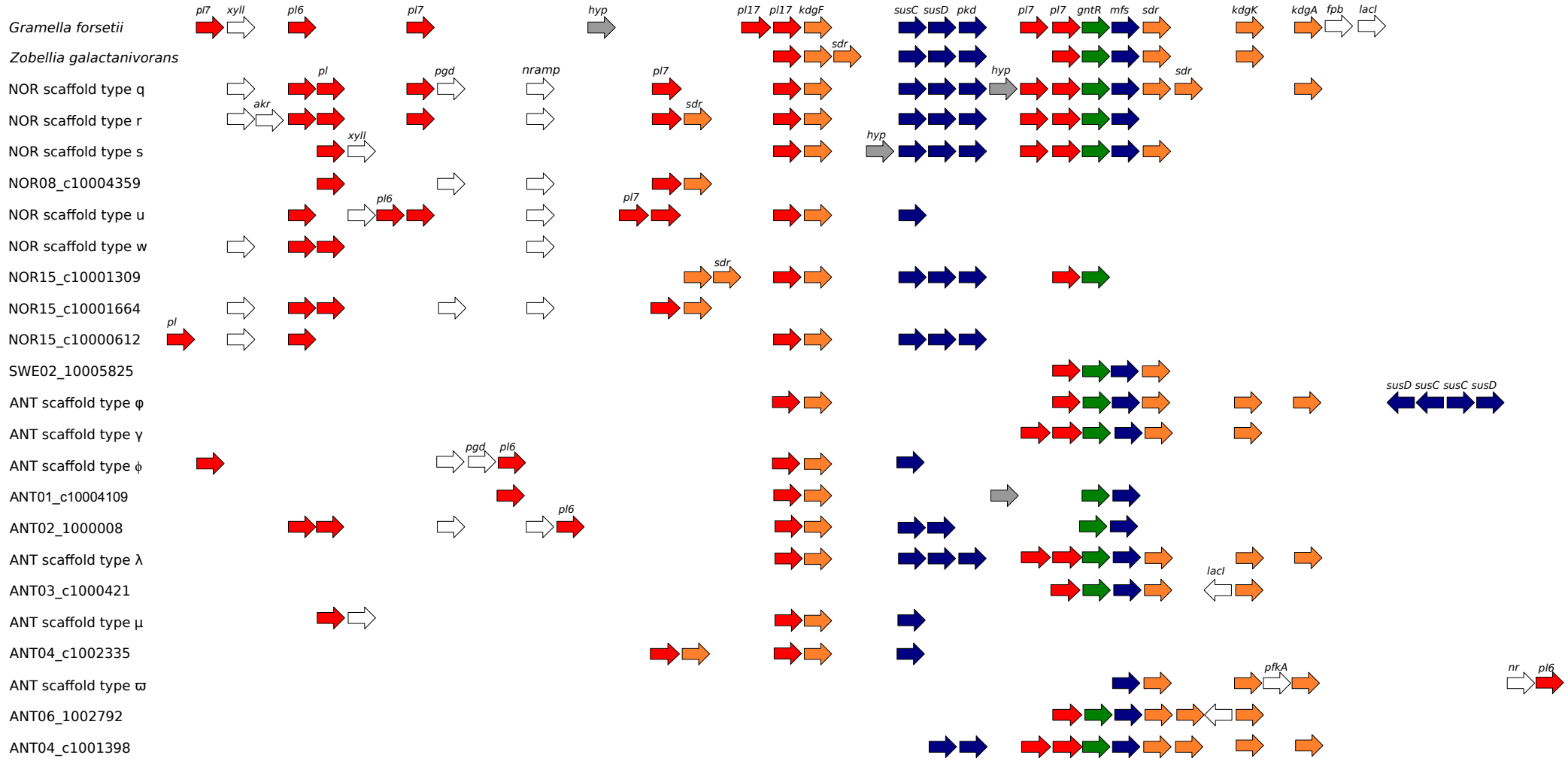

Figure S10. Alginolytic gene clusters in flavobacterial scaffolds and in genomes of flavobacterial strains. The scaffold number or the scaffold gene organization type (when more than one scaffold with the same gene organization were identified within the same sampling region) are indicated. The complete list of scaffolds can be found in Supplementary Table S5. Red, CDS coding for PL; green, trasnscriptional regulator; blue: transporters; gray, hypothetical proteins; orange, cytoplasmic processing enzymes; white, others. *pl*, polysaccharide lyase; *pl6*, *pl7* and *pl17*, AL or ALH belonging to the PL6, PL7 or PL17, respectively; *hyp*, hypothetical protein; *kdgF*, cupin domain; *susC*, outer membrane receptor protein; *susD*, SusD family; *pkd*, PKD-domain containing protein; *kdgK*, 2-dehydro-3-deoxygluconokinase; *kdgA*, 2-keto-3-deoxy-phosphogluconate aldolase; *gntR*, GntR family transcriptional regulator; *sdr*, short chain dehydrogenase/reductase superfamily; *mfs*, major facilitator superfamily transporter; *fpb*, fructose-1,6-biphosphatase; *lacI*, LacI family transcritptional regulator; *pgd*, 6-phosphogluconate dehydrogenase; *nramp*, Mn^{2+} and Fe^{2+} transporters of the NRAMP family; *akr*, Aldo/keto reductase, related to diketogulonate reductase; *xyll*, xylose isomerase-like TIM barrel.

References

Adami, M.L., Gordillo, S. (1999). Structure and dynamics of the biota associated with *Macrocystis pyrifera* (Phaeophyta) from the Beagle Channel, Tierra del Fuego. *Sci Mar* **63**: 183-191.

Alexandridis, N., Oschlies, A., and Wahl, M. (2012). Modeling the effects of abiotic and biotic factors on the depth distribution of *Fucus vesiculosus* in the Baltic Sea. *Mar Ecol-Prog Ser* **463**: 59-72.

Almanza, V., Buschmann, A.H. (2013). The ecological importance of *Macrocystis pyrifera* (Phaeophyta) forests towards a sustainable management and exploitation of Chilean coastal benthic co-management areas. *Int J Environ Sustain Develop* **12**: 341-360

Benkert, P., Tosatto, S.C., and Schomburg, D. (2008). QMEAN: A comprehensive scoring function for model quality assessment. *Proteins: Struct, Funct, Bioinf* **71**: 261-277.

Bergström, L., Tatarenkov, A., Johannesson, K., Jönsson, R.B., and Kautsky, L. (2005). Genetic and morphological identification of *Fucus radicans* sp. nov. (Fucales, Phaeophyceae) in the brackish Baltic Sea. *J Phycol* **41**: 1025-1038.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T. *et al.* (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res* **42**: W252-W258.

Bujalesky, G.G. (2007). Coastal geomorphology and evolution of Tierra del Fuego (Southern Argentina). *Geol Acta* **5**: 337-362.

Commendatore, M.G., Nievas, M.L., Amin, O., and Esteves, J.L. (2012). Sources and distribution of aliphatic and polyaromatic hydrocarbons in coastal sediments from the Ushuaia Bay (Tierra del Fuego, Patagonia, Argentina). *Mar Environ Res* **74**: 20-31.

Dauner, A.L.L., Hernández, E.A., MacCormack, W.P., and Martins, C.C. (2015). Molecular characterisation of anthropogenic sources of sedimentary organic matter from Potter Cove, King George Island, Antarctica. *Sci Total Environ* **502**: 408-416.

Dionisi, H.M., Lozada, M., Marcos, M.S., Di Marzio, W.D. and Loviso, C.L. (2011). Aromatic hydrocarbon degradation genes from chronically polluted Subantarctic marine sediments. In: De Bruijn, FJ (ed). Handbook of Molecular Microbial Ecology II: Metagenomics in Different Habitats. Wiley/Blackwell, pp 461-473.

Elmgren, R., Blenckner, T., and Andersson, A. (2015). Baltic Sea management: successes and failures. *Ambio* **44**: 335-344.

Ertesvåg, H., Erlien, F., Skjåk-Bræk, G., Rehm, B.H., and Valla, S. (1998). Biochemical properties and substrate specificities of a recombinantly produced *Azotobacter vinelandii* alginate lyase. *J Bacteriol* **180**: 3779-3784.

Evenset, A., G.N. Christensen & R. Palerud 2009. Environmental contaminants in marine sediments of Isfjorden, Svalbard 2009. Surveys outside Longyearbyen, Barentsburg, Pyramiden and Colesbukta. Akvaplan-niva report 4707-1 (in Norwegian).

Feller, G., and Gerday, C. (2003). Psychrophilic enzymes: hot topics in cold adaptation. *Nat Rev Microbiol* **1**: 200-208.

Garron, M,-L,, and Cygler, M. (2010). Structural and mechanistic classification of uronic acid-containing polysaccharide lyases. *Glycobiology* **20**: 1547-1573.

Garron, M.-L., and Cygler, M. (2014). Uronic polysaccharide degrading enzymes. *Curr Opin Struct Biol* **28**: 87-95.

Hansen, J., Sato, M., Russell, G., and Kharecha, P. (2013). Climate sensitivity, sea level and atmospheric carbon dioxide. *Philos T Roy Soc A* **371**: 20120294.

Holte, B., Dahle, S., Gulliksen, B., and Næs, K. (1996). Some macrofaunal effects of local pollution and glacier-induced sedimentation, with indicative chemical analyses, in the sediments of two Arctic fjords. *Polar Biol* **16**: 549-557.

Johannesson, K., Andre, C. (2006). Life on the margin: genetic isolation and diversity loss in a peripheral marine ecosystem, the Baltic Sea. *Mol Ecol* **15**: 2013-2029.

Korpinen, S., Meski, L., Andersen, J. H., Laamanen, M. (2012). Human pressures and their potential impact on the Baltic Sea ecosystem. *Ecol Indic* **15**: 105-114.

Krissinel, E. and Henrick, K. (2007). Protein interfaces, surfaces and assemblies' service PISA at the European Bioinformatics Institute. 'Inference of macromolecular assemblies from crystalline state'. *J Mol Biol* **372:** 774-797.

Krumhansl, K.A., Scheibling, R.E. (2012). Production and fate of kelp detritus. *Mar Ecol Prog Ser* **467**: 281-302.

Liuzzi, M.G., Gappa, J.L., and Piriz, M.L. (2011). Latitudinal gradients in macroalgal biodiversity in the Southwest Atlantic between 36 and 55°S. *Hydrobiologia* **673**: 205-214.

MacDonald, L.C., and Berger, B.W. (2014). Insight into the role of substrate-binding residues in conferring substrate specificity for the multifunctional polysaccharide lyase Smlt1473. *J Biol Chem* **289**: 18022-18032.

Markowitz, V.M., Chen, I.-M.A., Chu, K., Szeto, E., Palaniappan, K., Pillay, M. *et al.* (2014). IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucleic Acids Res* 42: D568-D573.

Mcwilliam, H., Valentin, F., Goujon, M., Li, W., Narayanasamy, M., Martin, J. *et al.* (2009). Web services at the European Bioinformatics Institute-2009. *Nucleic Acids Res* **37**: W6-W10.

Mikami, B., Ban, M., Suzuki, S., Yoon, H.-J., Miyake, O., Yamasaki, M. *et al.* (2011). Induced-fit motion of a lid loop involved in catalysis in alginate lyase A1-III. *Acta Crystallogr D* **68**: 1207-1216.

National Research Council (2003). Frontiers in Polar Biology in the Genomics Era. National Academies Press: Washington, DC, USA.

Nuth, C., Kohler, J., König, M., Deschwanden, A.V., Hagen, J.O.M., Kääb, A. *et al.* (2013). Decadal changes from a multi-temporal glacier inventory of Svalbard. *The Cryosphere* **7**: 1603-1621.

Patil, K.R., Roune, L., McHardy, A.C. (2012). The PhyloPythiaS web server for taxonomic assignment of metagenome sequences. *PLoS ONE* **7**: e38581.

Pereyra, R.T., Huenchuñir, C., Johansson, D., Forslund, H., Kautsky, L., Jonsson, P.R., Johannesson, K. (2013). Parallel speciation or long-distance dispersal? Lessons from seaweeds (*Fucus*) in the Baltic Sea. *J Evol Biol* **26**: 1727-1737.

Peters, A.F., Oppen, M.J., Wiencke, C., Stam, W.T., Olsen, J.L. (1997). Phylogeny and historical ecology of the Desmarestiaceae (Phaeophyceae) support a Southern Hemisphere origin. *J Phycol* **33**: 294-309.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. *et al.* (2004). UCSF Chimera-a visualization system for exploratory research and analysis. *J Comput Chem* **25**: 1605-1612.

Quartino, M., and De Zaixso, A.B. (2008). Summer macroalgal biomass in Potter Cove, South Shetland Islands, Antarctica: its production and flux to the ecosystem. *Polar Biol* **31**: 281-294.

Quartino, M.L., Deregibus, D., Campana, G.L., Latorre, G.E.J., and Momo, F.R. (2013). Evidence of macroalgal colonization on newly ice-free areas following glacial retreat in Potter Cove (South Shetland Islands), Antarctica. *PLoS ONE* **8**: e58223.

Reimann, S., Kallenborn, R., and Schmidbauer, N. (2009). Severe aromatic hydrocarbon pollution in the arctic town of Longyearbyen (Svalbard) caused by snowmobile emissions. *Environ Sci Technol* **43**: 4791-4795.

Renaud, P.E., Løkken, T.S., Jørgensen, L.L., Berge, J., Johnson, B.J. (2015). Macroalgal detritus and food-web subsidies along an Arctic fjord depth gradient. *Front Mar Sci* **2**: 31.

Rothäusler, E., Corell, H., Jormalainen, V. (2015). Abundance and dispersal trajectories of floating *Fucus vesiculosus* in the Northern Baltic Sea. *Limn Oceanog* **60**: 2173-2184.

Tatarek, A., Wiktor, J., Kendall, M.A. (2012). The sublittoral macroflora of Hornsund. *Polar Res* **31**: 18900.

Torn, K., Krause-Jensen, D., and Martin, G. (2006). Present and past depth distribution of bladderwrack (*Fucus vesiculosus*) in the Baltic Sea. *Aquat Bot* **84**: 53-62.

Vanella, F.A., Fernández, D.A., Romero, M. C., Calvo, J. (2007). Changes in the fish fauna associated with a sub-Antarctic *Macrocystis pyrifera* kelp forest in response to canopy removal. *Polar Biol* **30**: 449-457.

Velimirov, B., Field, J. G., Griffiths, C. L., and Zoutendyk, P. (1977). The ecology of kelp bed communities in the Benguela upwelling system. Analysis of biomass and spatial distribution. *Helgoländer Wiss Meeresuntersuch* **30**: 495–518.

Wiencke, C., and Amsler, C.D. (2012). Seaweeds and their communities in polar regions. *Seaweed Biology*. Springer. pp 265-291.

Wiencke, C.M., Clayton, N., Gómez, I., Iken, K., Lüder, U.H., Amsler, C.D., Karsten, U., Hanelt, D., Bischof, K., Dunton, K. (2007). Life strategy, ecophysiology and ecology of seaweeds in polar waters. *Rev Environ Sci Biotechnol* **6**: 95-126.

Willard, L., Ranjan, A., Zhang, H., Monzavi, H., Boyko, R.F., Syke,s B.D. *et al.* (2003). VADAR: a web server for quantitative evaluation of protein structure quality. *Nucleic Acids Res* **31**: 3316-3319.

Wulff, A., Iken, K., Quartino, M.L., Al-Handal, A., Wiencke, C., Clayton, M.N. (2009). Biodiversity, biogeography and zonation of marine benthic micro-and macroalgae in the Arctic and Antarctic. *Bot Mar* **52**: 491-507.

Yoon, H.-J., Hashimoto, W., Miyake, O., Murata, K., and Mikami, B. (2001). Crystal structure of alginate lyase A1-III complexed with trisaccharide product at 2.0 A resolution. *J Molec Biol* **307**: 9-16.